GenCore version 5.1.6

M protein - protein search, using sw model

un on: May 17, 2004, 12:46:47 ; Search time 41.5161 Seconds
(without alignments)
61.252 Million cell updates/sec

itle: US-09-458-299A-4233
erfect score: 43
equence: 1 KVFGSLAFV 9

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	43	100.0	9	4	AAB99689	Aab99689 HLA A2 bi
2	43	100.0	9	4	AAG88995	Aag88995 HER2/neu
3	43	100.0	9	4	AAG88786	Aag88786 HER2/neu
4	43	100.0	9	4	AAB75809	Aab75809 Tumour as
5	43	100.0	9	7	ADA49638	Ada49638 Multi-epi
6	43	100.0	144	7	ADA49445	Ada49445 Multi-epi
7	43	100.0	147	7	ADA49447	Ada49447 Multi-epi
8	43	100.0	148	7	ADA49443	Ada49443 Multi-epi
9	40	93.0	9	4	AAB99688	Aab99688 HLA A2 bi
10	40	93.0	9	4	AAG88994	Aag88994 HER2/neu
11	40	93.0	9	4	AAG88788	Aag88788 HER2/neu
12	40	93.0	9	5	AAU95942	Aau95942 Immunogen
13	40	93.0	9	5	AAU95940	Aau95940 Immunogen
14	39	90.7	9	2	AAR73685	Aar73685 Antigen f
15	39	90.7	9	2	AAR97507	Aar97507 Cytotoxic
16	39	90.7	9	2	AAW36824	Aaw36824 Immunogen
17	39	90.7	9	2	AAW70057	Aaw70057 HER-2/neu
18	39	90.7	9	2	AAW78859	Aaw78859 HER-1/neu
19	39	90.7	9	2	AAW77131	Aaw77131 HER-2/neu
20	39	90.7	9	2	AAY10495	Aay10495 HLA Class
21	39	90.7	9	3	AAB13755	Aab13755 Peptide f
22	39	90.7	9	3	AAB33671	Aab33671 MHC class
23	39	90.7	9	3	AAB23692	Aab23692 Cytotoxic
24	39	90.7	9	4	AAB74453	Aab74453 Her2/neu
25	39	90.7	9	4	AAB95917	Aab95917 MHC class
26	39	90.7	9	4	AAG93767	Aag93767 Human HER
27	39	90.7	9	4	AAB99696	Aab99696 HLA A2 bi
28	39	90.7	9	4	AAE00467	Aae00467 Human tum
29	39	90.7	9	4	AAG88785	Aag88785 HER2/neu
30	39	90.7	9	4	AAG88764	Aag88764 HER2/neu
31	39	90.7	9	4	AAG88993	Aag88993 HER2/neu
32	39	90.7	9	4	AAG88787	Aag88787 HER2/neu
33	39	90.7	9	4	AAG88996	Aag88996 HER2/neu
34	39	90.7	9	4	AAB68661	Aab68661 HER-2 CTL
35	39	90.7	9	4	AAU27089	Aau27089 Human Leu
36	39	90.7	9	4	AAU26756	Aau26756 Human Leu
37	39	90.7	9	4	AAB75810	Aab75810 Tumour as
38	39	90.7	9	5	ABG79077	Abg79077 Human HER
39	39	90.7	9	5	ABB76752	Abb76752 Tumour an
40	39	90.7	9	5	AAE26792	Aae26792 Human HLA
41	39	90.7	9	5	AAE19094	Aae19094 HLA-A2 re
42	39	90.7	9	5	ABG80178	Abg80178 MHC class
43	39	90.7	9	6	ABG72471	Abg72471 Cancer as
44	39	90.7	9	6	ABR56432	Abr56432 Her2/neu
45	39	90.7	9	6	ABR44533	Abr44533 Her2/neu

ALIGNMENTS

RESULT 1
AAB99689
ID AAB99689 standard; peptide; 9 AA.
XX
AC AAB99689;
XX
DT 06-SEP-2001 (first entry)
XX
DE HLA A2 binding CTL epitope peptide from Her2/neu SEQ ID NO:10.
XX
KW Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL;
KW cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu;
KW MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator;
KW immunotherapy; immune response.
XX
OS Homo sapiens.
XX
PN WO200141741-A1.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-US034318.
XX
PR 13-DEC-1999; 99US-0170448P.
PR 05-APR-2000; 2000US-00543608.
PR 30-MAY-2000; 2000US-00583200.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Celis E, Keogh E;
PI Chesnut R;
XX
DR WPI; 2001-381489/40.
XX
PT Compositions for use in a vaccine for treating, e.g., breast, lung and
PT colon cancer comprises at least one peptide that comprises an isolated
PT epitope of a tumor-associated antigen.
XX
PS Claim 1; Page 76; 86pp; English.
XX
CC The present invention describes a composition (I) comprising at least one
CC peptide that comprises an isolated, prepared epitope consisting of a
CC sequence selected from 25 short amino acid sequences given in AAB99680 to
CC AAB99704. Also described are: (1) a composition (II) comprising one or
CC more peptides, and further comprising at least two epitopes selected from
CC the 25 short amino acid sequences (as above), where each of the one or
CC more peptides comprise less than 50 contiguous amino acids that have 100%
CC identity with a native peptide sequence; and (2) a vaccine composition

```
C   (III) comprising an epitope selected from the 25 short amino acid
C   sequences (as above) and a pharmaceutical excipient. (I) has cytostatic
C   and immunomodulatory activities and can be used in vaccine production and
C   immunotherapy. The peptide epitope compositions (I)-(II) are useful for
C   monitoring an immune response to a tumour associated antigen or when one
C   or more peptides are combined to create a vaccine (III) that stimulates
C   the cellular arm of the immune system. In particular, the vaccine
C   mediates immune responses against tumours in individuals who bear an
C   allele of the human leukocyte antigen (HLA)-A2 supertype and improve the
C   standard of care for patients being treated for breast, colon, or lung
C   cancer
X
Q   Sequence 9 AA;

 Query Match             100.0%;  Score 43;  DB 4;  Length 9;
 Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
 Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

y           1 KVFGSLAFV 9
              |||||||||
b           1 KVFGSLAFV 9


ESULT 2
AG88995
D   AAG88995 standard; peptide; 9 AA.
X
C   AAG88995;
X
T   11-SEP-2001  (first entry)
X
E   HER2/neu epitope HLA-A2 supermotif-bearing peptide #8.
X
W   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
W   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
W   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
X
S   Homo sapiens.
S   Synthetic.
X
N   WO200141787-A1.
X
D   14-JUN-2001.
X
F   11-DEC-2000; 2000WO-US033591.
X
R   10-DEC-1999;   99US-00458299.
X
A   (EPIM-) EPIMMUNE INC.
X
I   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
I   Keogh E;
X
R   WPI; 2001-374995/39.
X
T   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
T   cellular immune responses for the prevention and treatment of cancer.
X
S   Claim 1; Page 189; 199pp; English.
X
C   The present invention describes isolated prepared HER2/neu epitopes (I).
C   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
C   culture in vitro and binds to a complex of an epitope (I), bound to a
C   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
C   and a second epitope and the peptide is less than 50 contiguous amino
C   acids that have 100% identity with a native peptide sequence of HER2/neu;
C   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
C   excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
C   ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
C   immunostimulant activities, and can be used in vaccines. (I), (II) and
C   (III) are useful for inducing cellular immune responses for the
C   prevention and treatment of cancer. (I) and (II) are useful for
C   monitoring or evaluating an immune response to a tumour-associated
CC  antigen when incubated with a T lymphocyte sample form a patient and
CC  detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC  based vaccines mean that immunosuppressive epitopes that may be present
CC  in whole antigens may be avoided. Selected epitopes may be combined to
CC  enhance immunogenicity. The possible pathological side effects caused by
CC  infectious agents or whole protein antigen is eliminated. The vaccine
CC  provides the ability to direct and focus an immune response to multiple
CC  selected antigens from the same pathogen. Epitope-based anti-tumour
CC  vaccines provides the opportunity to combine epitopes derived from
CC  multiple tumour-associated molecules addressing the problem of tumour-
CC  tumour variability and reducing the likelihood of tumour escape due to
CC  antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC  the exemplification of the present invention
XX
SQ  Sequence 9 AA;

  Query Match             100.0%;  Score 43;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |||||||||
Db          1 KVFGSLAFV 9


RESULT 3
AAG88786
ID  AAG88786 standard; peptide; 9 AA.
XX
AC  AAG88786;
XX
DT  11-SEP-2001  (first entry)
XX
DE  HER2/neu A2 supermotif crossbinding peptide #30.
XX
KW  Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW  immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW  tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS  Homo sapiens.
OS  Synthetic.
XX
PN  WO200141787-A1.
XX
PD  14-JUN-2001.
XX
PF  11-DEC-2000; 2000WO-US033591.
XX
PR  10-DEC-1999;   99US-00458299.
XX
PA  (EPIM-) EPIMMUNE INC.
XX
PI  Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI  Keogh E;
XX
DR  WPI; 2001-374995/39.
XX
PT  An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT  cellular immune responses for the prevention and treatment of cancer.
XX
PS  Example 2; Page 180; 199pp; English.
XX
CC  The present invention describes isolated prepared HER2/neu epitopes (I).
CC  Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC  culture in vitro and binds to a complex of an epitope (I), bound to a
CC  human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC  and a second epitope and the peptide is less than 50 contiguous amino
CC  acids that have 100% identity with a native peptide sequence of HER2/neu;
CC  (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC  excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC  ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC  immunostimulant activities, and can be used in vaccines. (I), (II) and
CC  (III) are useful for inducing cellular immune responses for the
```

prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample form a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour-tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in the exemplification of the present invention

Sequence 9 AA;

```
Query Match             100.0%;  Score 43;  DB 4;  Length 9;
Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

          1 KVFGSLAFV 9
            |||||||||
          1 KVFGSLAFV 9


SULT 4
B75809
```

AAB75809 standard; peptide; 9 AA.

AAB75809;

10-APR-2001 (first entry)

Tumour associated antigen Her2/neu HLA-A2 binding peptide.

Human leukocyte antigen; HLA; major histocompatibility complex; MHC; cytotoxic T lymphocyte; CTL; human class I MHC; immunogenic; HLA binding peptide; immune response; glycoprotein; cytostatic; virucide; hepatotropic; antiinflammatory; anti-HIV; vaccine; human immunodeficiency virus; protozoacide; viral infection; cancer; prostate cancer; hepatitis B; hepatitis C; human papilloma virus; HPV; cytomegalovirus; CMV; acquired immunodeficiency syndrome; AIDS; renal carcinoma; cervical carcinoma; lymphoma; malaria; condyloma acuminatum.

Homo sapiens.

WO200100225-A1.

04-JAN-2001.

28-JUN-2000; 2000WO-US017842.

29-JUN-1999; 99US-0141422P.

(EPIM-) EPIMMUNE INC.

Sette A, Sidney J, Southwood S;

WPI; 2001-112389/12.

Composition comprising human leukocyte antigen binding peptide which comprises isolated, prepared epitope useful for treating viral infections such as acquired immunodeficiency syndrome, and cancer.

Claim 1; Page 41; 58pp; English.

The present invention describes a composition (I) which comprises at least one human leukocyte antigen (HLA) binding peptide comprising an isolated, prepared epitope comprising one of 547 8-11 residue amino acid sequences (S1), given in AAB75803 to AAB76349. (I) has cytostatic, virucide, hepatotropic, antiinflammatory, anti-HIV (human immunodeficiency virus) and protozoacide activities, which can be used in vaccine production and is an inducer of cytotoxic T-cell response. (I) is useful for inducing a cytotoxic T cell response against a preselected antigen in a patient expressing a specific major histocompatibility complex (MHC) class I allele, by contacting cytotoxic T cells (CTLs) from the patient with (I). (I) is useful as a vaccine to treat and/or prevent viral infection and cancer such as prostate cancer, hepatitis B, hepatitis C, human papilloma virus (HPV) infection, cytomegalovirus (CMV), acquired immunodeficiency syndrome (AIDS), renal carcinoma, cervical carcinoma, lymphoma, malaria, and condyloma acuminatum

XX

SQ Sequence 9 AA;

```
  Query Match             100.0%;  Score 43;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |||||||||
Db          1 KVFGSLAFV 9


RESULT 5
ADA49638
```

ID ADA49638 standard; peptide; 9 AA.

XX

AC ADA49638;

XX

DT 20-NOV-2003 (first entry)

XX

DE Multi-epitope construct specific epitope #180.

XX

KW multi-epitope; immunogenic; epitope; major histocompatibility complex; MHC class I; MHC class II; junctional epitope.

XX

OS Unidentified.

XX

PN US2002119127-A1.

XX

PD 29-AUG-2002.

XX

PF 27-JUN-2001; 2001US-00894018.

XX

PR 28-DEC-1999; 99US-0173390P.
PR 28-DEC-2000; 2000WO-US035568.
PR 16-APR-2001; 2001US-0284221P.

XX

PA (SETT/) SETTE A.
PA (CHES/) CHESNUT R.
PA (LIVI/) LIVINGSTON B D.
PA (BAKE/) BAKER D M.
PA (NEWM/) NEWMAN M J.
PA (BROW/) BROWN D H.

XX

PI Sette A, Chesnut R, Livingston BD, Baker DM, Newman MJ, Brown DH;

XX

DR WPI; 2003-615704/58.

XX

PT Designing multi-epitope construct having major histocompatability complex
PT class I and II epitope nucleic acids, by selecting mixture of amino acid
PT insertions at junctions of construct to minimize junctional epitopes.

XX

PS Disclosure; Fig 19E; 78pp; English.

XX

CC The invention relates to a method of designing multi-epitope constructs
CC comprising major histocompatibility complex (MHC) class I and II (CTL)
CC epitope nucleic acids (CEN), involves sorting CEN, introducing flanking
CC amino acid residue selected from specified amino acid residues given in
CC specification at C+1 position of CEN, introducing amino acid spacer
CC residues between two CEN, and selecting the constructs having less
CC junctional epitopes. The method is useful for designing a multi-epitope
CC construct having multiple epitope nucleic acid. The method avoids or

CC minimises the occurrence of junctional epitopes and maximises the
CC immunogenicity and/or antigenicity of multi-epitope vaccines. The present
CC sequence represents the amino acid sequence of an epitope present in a
CC multi-epitope construct.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 43; DB 7; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
||||||||| |
Db 1 KVFGSLAFV 9

RESULT 6
ADA49445
ID ADA49445 standard; protein; 144 AA.
XX
AC ADA49445;
XX
DT 20-NOV-2003 (first entry)
XX
DE Multi-epitope construct #25.
XX
KW multi-epitope; immunogenic; epitope; major histocompatibility complex;
KW MHC class I; MHC class II; junctional epitope.
XX
OS Synthetic.
XX
PN US2002119127-A1.
XX
PD 29-AUG-2002.
XX
PF 27-JUN-2001; 2001US-00894018.
XX
PR 28-DEC-1999; 99US-0173390P.
PR 28-DEC-2000; 2000WO-US035568.
PR 16-APR-2001; 2001US-0284221P.
XX
PA (SETT/) SETTE A.
PA (CHES/) CHESNUT R.
PA (LIVI/) LIVINGSTON B D.
PA (BAKE/) BAKER D M.
PA (NEWM/) NEWMAN M J.
PA (BROW/) BROWN D H.
XX
PI Sette A, Chesnut R, Livingston BD, Baker DM, Newman MJ, Brown DH;
XX
DR WPI; 2003-615704/58.
DR N-PSDB; ADA49446.
XX
PT Designing multi-epitope construct having major histocompatability complex
PT class I and II epitope nucleic acids, by selecting mixture of amino acid
PT insertions at junctions of construct to minimize junctional epitopes.
XX
PS Disclosure; Fig 18K; 78pp; English.
XX
CC The invention relates to a method of designing multi-epitope constructs
CC comprising major histocompatibility complex (MHC) class I and II (CTL)
CC epitope nucleic acids (CEN), involves sorting CEN, introducing flanking
CC amino acid residue selected from specified amino acid residues given in
CC specification at C+1 position of CEN, introducing amino acid spacer
CC residues between two CEN, and selecting the constructs having less
CC junctional epitopes. The method is useful for designing a multi-epitope
CC construct having multiple epitope nucleic acid. The method avoids or
CC minimises the occurrence of junctional epitopes and maximises the
CC immunogenicity and/or antigenicity of multi-epitope vaccines. The present
CC sequence represents the amino acid sequence of a multi-epitope construct.
XX
SQ Sequence 144 AA;

Query Match 100.0%; Score 43; DB 7; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
||||||||| |
Db 102 KVFGSLAFV 110

RESULT 7
ADA49447
ID ADA49447 standard; protein; 147 AA.
XX
AC ADA49447;
XX
DT 20-NOV-2003 (first entry)
XX
DE Multi-epitope construct #26.
XX
KW multi-epitope; immunogenic; epitope; major histocompatibility complex;
KW MHC class I; MHC class II; junctional epitope.
XX
OS Synthetic.
XX
PN US2002119127-A1.
XX
PD 29-AUG-2002.
XX
PF 27-JUN-2001; 2001US-00894018.
XX
PR 28-DEC-1999; 99US-0173390P.
PR 28-DEC-2000; 2000WO-US035568.
PR 16-APR-2001; 2001US-0284221P.
XX
PA (SETT/) SETTE A.
PA (CHES/) CHESNUT R.
PA (LIVI/) LIVINGSTON B D.
PA (BAKE/) BAKER D M.
PA (NEWM/) NEWMAN M J.
PA (BROW/) BROWN D H.
XX
PI Sette A, Chesnut R, Livingston BD, Baker DM, Newman MJ, Brown DH;
XX
DR WPI; 2003-615704/58.
DR N-PSDB; ADA49448.
XX
PT Designing multi-epitope construct having major histocompatability complex
PT class I and II epitope nucleic acids, by selecting mixture of amino acid
PT insertions at junctions of construct to minimize junctional epitopes.
XX
PS Disclosure; Fig 18K; 78pp; English.
XX
CC The invention relates to a method of designing multi-epitope constructs
CC comprising major histocompatibility complex (MHC) class I and II (CTL)
CC epitope nucleic acids (CEN), involves sorting CEN, introducing flanking
CC amino acid residue selected from specified amino acid residues given in
CC specification at C+1 position of CEN, introducing amino acid spacer
CC residues between two CEN, and selecting the constructs having less
CC junctional epitopes. The method is useful for designing a multi-epitope
CC construct having multiple epitope nucleic acid. The method avoids or
CC minimises the occurrence of junctional epitopes and maximises the
CC immunogenicity and/or antigenicity of multi-epitope vaccines. The present
CC sequence represents the amino acid sequence of a multi-epitope construct.
XX
SQ Sequence 147 AA;

Query Match 100.0%; Score 43; DB 7; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
||||||||| |
Db 76 KVFGSLAFV 84

SULT 8
A49443
ADA49443 standard; protein; 148 AA.

ADA49443;

20-NOV-2003 (first entry)

Multi-epitope construct #24.

multi-epitope; immunogenic; epitope; major histocompatibility complex; MHC class I; MHC class II; junctional epitope.

Synthetic.

US2002119127-A1.

29-AUG-2002.

27-JUN-2001; 2001US-00894018.

28-DEC-1999; 99US-0173390P.
28-DEC-2000; 2000WO-US035568.
16-APR-2001; 2001US-0284221P.

(SETT/) SETTE A.
(CHES/) CHESNUT R.
(LIVI/) LIVINGSTON B D.
(BAKE/) BAKER D M.
(NEWM/) NEWMAN M J.
(BROW/) BROWN D H.

Sette A, Chesnut R, Livingston BD, Baker DM, Newman MJ, Brown DH;

WPI; 2003-615704/58.
N-PSDB; ADA49444.

Designing multi-epitope construct having major histocompatability complex class I and II epitope nucleic acids, by selecting mixture of amino acid insertions at junctions of construct to minimize junctional epitopes.

Disclosure; Fig 18K; 78pp; English.

The invention relates to a method of designing multi-epitope constructs comprising major histocompatibility complex (MHC) class I and II (CTL) epitope nucleic acids (CEN), involves sorting CEN, introducing flanking amino acid residue selected from specified amino acid residues given in specification at C+1 position of CEN, introducing amino acid spacer residues between two CEN, and selecting the constructs having less junctional epitopes. The method is useful for designing a multi-epitope construct having multiple epitope nucleic acid. The method avoids or minimises the occurrence of junctional epitopes and maximises the immunogenicity and/or antigenicity of multi-epitope vaccines. The present sequence represents the amino acid sequence of a multi-epitope construct.

Sequence 148 AA;

Query Match 100.0%; Score 43; DB 7; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
|||||||||
Db 53 KVFGSLAFV 61

ESULT 9
AB99688
AAB99688 standard; peptide; 9 AA.

AAB99688;

XX
DT 06-SEP-2001 (first entry)
XX
DE HLA A2 binding CTL epitope peptide from Her2/neu SEQ ID NO:9.
XX
KW Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL;
KW cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu;
KW MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator;
KW immunotherapy; immune response.
XX
OS Homo sapiens.
XX
PN WO200141741-A1.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-US034318.
XX
PR 13-DEC-1999; 99US-0170448P.
PR 05-APR-2000; 2000US-00543608.
PR 30-MAY-2000; 2000US-00583200.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Celis E, Keogh E;
PI Chesnut R;
XX
DR WPI; 2001-381489/40.
XX
PT Compositions for use in a vaccine for treating, e.g., breast, lung and
PT colon cancer comprises at least one peptide that comprises an isolated
PT epitope of a tumor-associated antigen.
XX
PS Claim 1; Page 76; 86pp; English.
XX
CC The present invention describes a composition (I) comprising at least one
CC peptide that comprises an isolated, prepared epitope consisting of a
CC sequence selected from 25 short amino acid sequences given in AAB99680 to
CC AAB99704. Also described are: (1) a composition (II) comprising one or
CC more peptides, and further comprising at least two epitopes selected from
CC the 25 short amino acid sequences (as above), where each of the one or
CC more peptides comprise less than 50 contiguous amino acids that have 100%
CC identity with a native peptide sequence; and (2) a vaccine composition
CC (III) comprising an epitope selected from the 25 short amino acid
CC sequences (as above) and a pharmaceutical excipient. (I) has cytostatic
CC and immunomodulatory activities and can be used in vaccine production and
CC immunotherapy. The peptide epitope compositions (I)-(II) are useful for
CC monitoring an immune response to a tumour associated antigen or when one
CC or more peptides are combined to create a vaccine (III) that stimulates
CC the cellular arm of the immune system. In particular, the vaccine
CC mediates immune responses against tumours in individuals who bear an
CC allele of the human leukocyte antigen (HLA)-A2 supertype and improve the
CC standard of care for patients being treated for breast, colon, or lung
CC cancer
XX
SQ Sequence 9 AA;

Query Match 93.0%; Score 40; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
|:||||||
Db 1 KLFGSLAFV 9

RESULT 10
AAG88994
ID AAG88994 standard; peptide; 9 AA.
XX
AC AAG88994;
XX
DT 11-SEP-2001 (first entry)

```
X
E   HER2/neu epitope HLA-A2 supermotif-bearing peptide #7.
X
W   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
W   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
W   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
X
S   Homo sapiens.
S   Synthetic.
X
N   WO200141787-A1.
X
D   14-JUN-2001.
X
F   11-DEC-2000; 2000WO-US033591.
X
R   10-DEC-1999;   99US-00458299.
X
A   (EPIM-) EPIMMUNE INC.
X
I   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
I   Keogh E;
X
R   WPI; 2001-374995/39.
X
T   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
T   cellular immune responses for the prevention and treatment of cancer.
X
S   Claim 1; Page 189; 199pp; English.
X
C   The present invention describes isolated prepared HER2/neu epitopes (I).
C   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
C   culture in vitro and binds to a complex of an epitope (I), bound to a
C   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
C   and a second epitope and the peptide is less than 50 contiguous amino
C   acids that have 100% identity with a native peptide sequence of HER2/neu;
C   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
C   excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
C   ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
C   immunostimulant activities, and can be used in vaccines. (I), (II) and
C   (III) are useful for inducing cellular immune responses for the
C   prevention and treatment of cancer. (I) and (II) are useful for
C   monitoring or evaluating an immune response to a tumour-associated
C   antigen when incubated with a T lymphocyte sample form a patient and
C   detecting the presence of bound T lymphocyte to (I) or (II). Epitope
C   based vaccines mean that immunosuppressive epitopes that may be present
C   in whole antigens may be avoided. Selected epitopes may be combined to
C   enhance immunogenicity. The possible pathological side effects caused by
C   infectious agents or whole protein antigen is eliminated. The vaccine
C   provides the ability to direct and focus an immune response to multiple
C   selected antigens from the same pathogen. Epitope-based anti-tumour
C   vaccines provides the opportunity to combine epitopes derived from
C   multiple tumour-associated molecules addressing the problem of tumour-
C   tumour variability and reducing the likelihood of tumour escape due to
C   antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
C   the exemplification of the present invention
X
Q   Sequence 9 AA;

 Query Match             93.0%;  Score 40;  DB 4;  Length 9;
 Best Local Similarity   88.9%;  Pred. No. 1.4e+06;
 Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

y           1 KVFGSLAFV 9
              |:|||||||
b           1 KLFGSLAFV 9


ESULT 11
AG88788
D   AAG88788 standard; peptide; 9 AA.
X
C   AAG88788;
```

```
XX
DT   11-SEP-2001  (first entry)
XX
DE   HER2/neu A2 supermotif crossbinding peptide #32.
XX
KW   Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW   immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW   tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200141787-A1.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US033591.
XX
PR   10-DEC-1999;   99US-00458299.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI   Keogh E;
XX
DR   WPI; 2001-374995/39.
XX
PT   An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT   cellular immune responses for the prevention and treatment of cancer.
XX
PS   Example 2; Page 180; 199pp; English.
XX
CC   The present invention describes isolated prepared HER2/neu epitopes (I).
CC   Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC   culture in vitro and binds to a complex of an epitope (I), bound to a
CC   human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC   and a second epitope and the peptide is less than 50 contiguous amino
CC   acids that have 100% identity with a native peptide sequence of HER2/neu;
CC   (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC   excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC   ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC   immunostimulant activities, and can be used in vaccines. (I), (II) and
CC   (III) are useful for inducing cellular immune responses for the
CC   prevention and treatment of cancer. (I) and (II) are useful for
CC   monitoring or evaluating an immune response to a tumour-associated
CC   antigen when incubated with a T lymphocyte sample form a patient and
CC   detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC   based vaccines mean that immunosuppressive epitopes that may be present
CC   in whole antigens may be avoided. Selected epitopes may be combined to
CC   enhance immunogenicity. The possible pathological side effects caused by
CC   infectious agents or whole protein antigen is eliminated. The vaccine
CC   provides the ability to direct and focus an immune response to multiple
CC   selected antigens from the same pathogen. Epitope-based anti-tumour
CC   vaccines provides the opportunity to combine epitopes derived from
CC   multiple tumour-associated molecules addressing the problem of tumour-
CC   tumour variability and reducing the likelihood of tumour escape due to
CC   antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC   the exemplification of the present invention
XX
SQ   Sequence 9 AA;

  Query Match             93.0%;  Score 40;  DB 4;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:|||||||
Db          1 KLFGSLAFV 9


RESULT 12
AAU95942
ID   AAU95942 standard; peptide; 9 AA.
```

AAU95942;

02-JUL-2002 (first entry)

Immunogenic peptide with (HLA)-A2.1 binding site #155.

HLA-A2.1 binding peptide; cytostatic; virucide; anti-HIV; hepatotropic;
human immunodeficiency virus; antiinflammatory; antibacterial; vaccine;
protozoacide; immunosuppressant; immunogenic peptide; T cell activation;
human leucocyte antigen binding site; cytotoxic T cell response;
viral infection; hepatitis; Epstein-Barr virus; papilloma virus;
human immunodeficiency virus; HIV; Kaposi sarcoma; Lassa fever virus;
cytomegalovirus; tumour; prostate cancer; renal carcinoma; lymphona;
prostate-specific antigen; p53; carcino-embryonal antigen;
melanoma antigen; Mycobacterium tuberculosis; protozoa;
trypanosome surface antigen; condlyloma acuminatum.

Unidentified.

WO200220616-A1.

14-MAR-2002.

01-SEP-2000; 2000WO-US024102.

01-SEP-2000; 2000WO-US024102.

(EPIM-) EPIMMUNE INC.

Grey HM, Sette A, Sidney J, Southwood S;

WPI; 2002-351766/38.

Immunogenic peptide with human leucocyte antigen-A2.1 binding site,
useful for treating e.g. viral infection or tumors.

Claim 1; Page 29; 35pp; English.

The invention describes a composition comprising an immunogenic peptide
having a human leucocyte antigen (HLA)-A2.1 binding site. The peptides
bind specifically to HLA-A2.1, to cause T cell activation and thus a
cytotoxic T cell response. The peptides and the nucleic acids that
encodes them, are used, in vivo or ex vivo, for treatment of viral
infections (hepatitis B or C; Epstein-Barr; human immune deficiency;
Kaposi sarcoma; human papilloma; Lassa fever or cytomegaloviruses);
tumours including prostate cancer, renal carcinoma and lymphona (where
directed to prostate-specific antigen, p53, carcino-embryonal antigen,
Her2/neu or melanoma antigens); infection by Mycobacterium tuberculosis
or protozoa (directed to trypanosome surface antigen); and condlyloma
acuminatum. The peptides are suitable for use in peptide-based vaccines.
This sequence represents an immunogenic peptide with the human leucocyte
antigen (HLA)-A2.1 binding site, described in the invention

Sequence 9 AA;

```
Query Match             93.0%;  Score 40;  DB 5;  Length 9;
Best Local Similarity   88.9%;  Pred. No. 1.4e+06;
Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:|||||||
Db          1 KMFGSLAFV 9
```

RESULT 13
AAU95940
AAU95940 standard; peptide; 9 AA.

AAU95940;

02-JUL-2002 (first entry)

Immunogenic peptide with (HLA)-A2.1 binding site #153.

HLA-A2.1 binding peptide; cytostatic; virucide; anti-HIV; hepatotropic;
human immunodeficiency virus; antiinflammatory; antibacterial; vaccine;
protozoacide; immunosuppressant; immunogenic peptide; T cell activation;
human leucocyte antigen binding site; cytotoxic T cell response;
viral infection; hepatitis; Epstein-Barr virus; papilloma virus;
human immunodeficiency virus; HIV; Kaposi sarcoma; Lassa fever virus;
cytomegalovirus; tumour; prostate cancer; renal carcinoma; lymphona;
prostate-specific antigen; p53; carcino-embryonal antigen;
melanoma antigen; Mycobacterium tuberculosis; protozoa;
trypanosome surface antigen; condlyloma acuminatum.

Unidentified.

WO200220616-A1.

14-MAR-2002.

01-SEP-2000; 2000WO-US024102.

01-SEP-2000; 2000WO-US024102.

(EPIM-) EPIMMUNE INC.

Grey HM, Sette A, Sidney J, Southwood S;

WPI; 2002-351766/38.

Immunogenic peptide with human leucocyte antigen-A2.1 binding site,
useful for treating e.g. viral infection or tumors.

Claim 1; Page 29; 35pp; English.

The invention describes a composition comprising an immunogenic peptide
having a human leucocyte antigen (HLA)-A2.1 binding site. The peptides
bind specifically to HLA-A2.1, to cause T cell activation and thus a
cytotoxic T cell response. The peptides and the nucleic acids that
encodes them, are used, in vivo or ex vivo, for treatment of viral
infections (hepatitis B or C; Epstein-Barr; human immune deficiency;
Kaposi sarcoma; human papilloma; Lassa fever or cytomegaloviruses);
tumours including prostate cancer, renal carcinoma and lymphona (where
directed to prostate-specific antigen, p53, carcino-embryonal antigen,
Her2/neu or melanoma antigens); infection by Mycobacterium tuberculosis
or protozoa (directed to trypanosome surface antigen); and condlyloma
acuminatum. The peptides are suitable for use in peptide-based vaccines.
This sequence represents an immunogenic peptide with the human leucocyte
antigen (HLA)-A2.1 binding site, described in the invention

Sequence 9 AA;

```
Query Match             93.0%;  Score 40;  DB 5;  Length 9;
Best Local Similarity   88.9%;  Pred. No. 1.4e+06;
Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:|||||||
Db          1 KLFGSLAFV 9
```

RESULT 14
AAR73685
AAR73685 standard; peptide; 9 AA.

AAR73685;

25-MAR-2003 (revised)
14-JUN-1995 (first entry)

Antigen fragment 1, from c-ERB2 has binding affinity for HLA-2.1.

antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV; HIV1;
plasma specific antigen; hepatitis B virus; Epstein Barr;

```
W   human immunodeficiency virus; human papilloma virus; p53; c-ERB2; MAGE-1;
W   melanoma antigen-1; core antigen; surface antigen;
W   pharmaceutical composition; in vivo; ex vivo; therapeutic; diagnostic;
W   MHC class I molecule; major histocompatability complex; HLA-A2.1; 9mer;
W   10mer; anchor; human leukocyte antigen; PLP; 8mer; algorithm prediction;
W   MBP; CMV; cytomegalovirus; HSV; herpes simplex virus.
X
S   Homo sapiens.
X
N   WO9420127-A1.
X
D   15-SEP-1994.
X
F   04-MAR-1994;   94WO-US002353.
X
R   05-MAR-1993;   93US-00027146.
R   04-JUN-1993;   93US-00073205.
R   29-NOV-1993;   93US-00159184.
X
A   (CYTE-) CYTEL CORP.
X
I   Grey HM,  Sette A,  Sidney J,  Kast W;
X
R   WPI; 1994-302678/37.
X
T   Immunogenic peptide(s) having an HLA-A2.1 binding motif - used for
T   treatment or prophylaxis of cancer, virus infection or autoimmune
T   diseases.
X
S   Disclosure; Page 80; 138pp; English.
X
C   AAR73685-876 are potential peptide binders of HLA-A2.1 motif. Using
C   motifs disclosed in the invention, these peptides were screened for
C   further motifs. Only peptides with binding affinity of at least 1%
C   (binding affinity is expressed as an IC50 value) as compared to the
C   standard peptide (AAR71293) in assays. This peptide has an binding value
C   of 0.1500. The peptides of the invention can induce cytotoxic T
C   lymphocytes which can react with target cells. They can be used for the
C   treatment or prophlyaxis of cancer, eg. prostate cancer or lymphoma, etc.
C   (Updated on 25-MAR-2003 to correct PN field.)
X
Q   Sequence 9 AA;

 Query Match             90.7%;  Score 39;  DB 2;  Length 9;
 Best Local Similarity   77.8%;  Pred. No. 1.4e+06;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Y           1 KVFGSLAFV 9
              |:||||||:
b           1 KIFGSLAFL 9


ESULT 15
AR97507
D   AAR97507 standard; peptide; 9 AA.
X
C   AAR97507;
X
T   11-FEB-1997  (first entry)
X
E   Cytotoxic T lymphocyte-activating Her-2/Neu-specific peptide.
X
W   p53; Her-2; Neu; aa; amino acid; CTL; cytotoxic T lymphocyte; target;
W   malignant cell; antigenic; vaccine; immunisation; activation.
X
S   Homo sapiens.
X
N   WO9618409-A1.
X
D   20-JUN-1996.
X
F   14-DEC-1995;   95WO-US016415.
X
PR  14-DEC-1994;   94US-00355558.
XX
PA  (SCRI ) SCRIPPS RES INST.
XX
PI  Sherman LA;
XX
DR  WPI; 1996-300385/30.
XX
PT  In vivo activation of tumour-specific cytotoxic T lymphocytes - by
PT  contacting with polypeptide(s) derived from human p53 or Her-2/Neu
PT  proteins.
XX
PS  Claim 5; Page 124; 158pp; English.
XX
CC  AAR97507 is a peptide capable of activating cytotoxic T lymphocytes
CC  (CTLs) which specifically target malignant cells. The peptide corresponds
CC  to amino acids 369-377 of human Her-2/Neu protein. CTL- activating
CC  peptides can be used in a vaccine for protecting against tumour cell
CC  formation. CTLs activated by the peptides will lyse tumour cells
CC  displaying specific peptides. Antibodies against CTL- activating peptides
CC  are useful for the identification of other similar compounds which may be
CC  useful for treating cancer or virally- infected cells, or for diagnosis.
CC  The peptide and vaccines produced provide immunity to a high percentage
CC  of different ethnic groups, i.e. those with different HLA alleles
XX
SQ  Sequence 9 AA;

  Query Match             90.7%;  Score 39;  DB 2;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 1.4e+06;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:||||||:
Db          1 KIFGSLAFL 9


Search completed: May 17, 2004, 12:54:32
Job time : 42.5161 secs
```

GenCore version 5.1.6

M protein - protein search, using sw model

un on: May 17, 2004, 12:47:22 ; Search time 28.7419 Seconds
(without alignments)
98.799 Million cell updates/sec

itle: US-09-458-299A-4233
erfect score: 43
equence: 1 KVFGSLAFV 9

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 1017041 seqs, 315518202 residues

otal number of hits satisfying chosen parameters: 1017041

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	39	90.7	711	11	Q80Y89	Q80y89 mus musculu
2	39	90.7	1259	6	O18735	O18735 canis famil
3	39	90.7	1259	11	Q8K3F9	Q8k3f9 rattus norv
4	38	88.4	1294	10	Q8LGU1	Q8lgu1 arabidopsis
5	38	88.4	1306	10	Q9LU34	Q9lu34 arabidopsis
6	36	83.7	333	5	Q9GUC0	Q9guc0 caenorhabdi
7	36	83.7	431	13	Q7SY19	Q7sy19 brachydanio
8	35	81.4	384	16	Q7VNU0	Q7vnu0 haemophilus
9	35	81.4	1708	10	Q7XWZ9	Q7xwz9 oryza sativ
10	35	81.4	4180	16	Q9I5N6	Q9i5n6 pseudomonas
11	34	79.1	43	16	Q81VA0	Q81va0 bacillus an
12	34	79.1	59	9	Q8LTR7	Q8ltr7 lactococcus
13	34	79.1	59	9	Q8LTL8	Q8ltl8 lactococcus
14	34	79.1	133	16	Q83CR5	Q83cr5 coxiella bu
15	34	79.1	134	2	Q8GLF9	Q8glf9 streptococc
16	34	79.1	181	16	Q8FXS3	Q8fxs3 brucella su
17	34	79.1	181	16	Q82N79	Q82n79 streptomyce
18	34	79.1	249	16	Q8D479	Q8d479 vibrio vuln
19	34	79.1	454	5	Q9XTU7	Q9xtu7 caenorhabdi
20	34	79.1	553	5	Q8SUF9	Q8suf9 encephalito
21	34	79.1	562	13	Q7T2C0	Q7t2c0 brachydanio
22	34	79.1	769	16	Q7UJS5	Q7ujs5 rhodopirell
23	34	79.1	822	10	Q9SHL9	Q9shl9 arabidopsis
24	34	79.1	3371	12	Q9J9C2	Q9j9c2 apoi virus.
25	33	76.7	231	5	Q19158	Q19158 caenorhabdi
26	33	76.7	297	16	Q9Z8S3	Q9z8s3 chlamydia p
27	33	76.7	331	2	Q8VM84	Q8vm84 rhizobium l
28	33	76.7	335	5	O17809	O17809 caenorhabdi
29	33	76.7	343	5	Q9UAX3	Q9uax3 caenorhabdi
30	33	76.7	370	10	Q9LYU5	Q9lyu5 arabidopsis
31	33	76.7	391	17	Q979S3	Q979s3 thermoplasm
32	33	76.7	404	10	Q9FFX8	Q9ffx8 arabidopsis
33	33	76.7	443	16	Q8A8Y3	Q8a8y3 bacteroides
34	33	76.7	505	5	Q9U2K5	Q9u2k5 caenorhabdi
35	33	76.7	602	16	Q8KDR8	Q8kdr8 chlorobium
36	32	74.4	96	17	Q8TQN9	Q8tqn9 methanosarc
37	32	74.4	105	8	Q85NZ2	Q85nz2 naubates fu
38	32	74.4	125	8	Q8M3S3	Q8m3s3 naubates fu
39	32	74.4	125	8	Q8LVT8	Q8lvt8 naubates ha
40	32	74.4	129	16	Q8Z1I5	Q8z1i5 salmonella
41	32	74.4	180	16	Q9EX23	Q9ex23 streptomyce
42	32	74.4	244	5	Q86FB4	Q86fb4 schistosoma
43	32	74.4	257	8	Q8WA73	Q8wa73 bemisia tab
44	32	74.4	337	16	Q89P67	Q89p67 bradyrhizob
45	32	74.4	342	16	Q98FZ0	Q98fz0 rhizobium l

ALIGNMENTS

RESULT 1
Q80Y89
ID Q80Y89 PRELIMINARY; PRT; 711 AA.
AC Q80Y89;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE V-erb-b2 erythroblastic leukemia viral oncogene homolog 2,
DE neuro/glioblastoma derived oncogene homolog (Hypothetical
DE protein).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC046811; AAH46811.1; -.
DR   EMBL; BC053078; AAH53078.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0005006; F:epidermal growth factor receptor activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; IEA.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR006211; Furin-like.
DR   InterPro; IPR006212; Furin_repeat.
DR   InterPro; IPR009030; Grow_fac_recep.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   SMART; SM00261; FU; 4.
GW   Hypothetical protein.
SQ   SEQUENCE   711 AA;  78707 MW;  682B188EB0E71318 CRC64;

 Query Match             90.7%;  Score 39;  DB 11;  Length 711;
 Best Local Similarity   77.8%;  Pred. No. 13;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:||||||:
Db        370 KIFGSLAFL 378


RESULT 2
O18735
ID   O18735      PRELIMINARY;      PRT;  1259 AA.
AC   O18735;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ErbB-2.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yokota H.;
RT   "cDNA cloning of erbB-2 from canine mammary gland.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB008451; BAA23127.1; -.
DR   HSSP; P11362; 1FGK.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005006; F:epidermal growth factor receptor activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; IEA.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR006211; Furin-like.
DR   InterPro; IPR006212; Furin_repeat.
DR   InterPro; IPR009030; Grow_fac_recep.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00261; FU; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ   SEQUENCE   1259 AA;  137989 MW;  E37364D49C4ACD46 CRC64;

  Query Match             90.7%;  Score 39;  DB 6;  Length 1259;
  Best Local Similarity   77.8%;  Pred. No. 23;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:||||||:
Db        369 KIFGSLAFL 377


RESULT 3
Q8K3F9
ID   Q8K3F9      PRELIMINARY;      PRT;  1259 AA.
AC   Q8K3F9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neu protooncoprotein.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BDIX;
RA   Watson P.A., Kim K., Chen K.-S., Gould M.N.;
RT   "Androgen-Dependent Mammary Carcinogenesis in Rats Transgenic for the
RT   Neu Proto-Oncogene.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY116182; AAM50093.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005006; F:epidermal growth factor receptor activity; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; IEA.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR006211; Furin-like.
DR   InterPro; IPR009030; Grow_fac_recep.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF01030; Recep_L_domain; 2.
DR   Pfam; PF02757; YLP; 2.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   1259 AA;  139102 MW;  B724BD5CC33AE953 CRC64;
```

```
 Query Match             90.7%;  Score 39;  DB 11;  Length 1259;
 Best Local Similarity   77.8%;  Pred. No. 23;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:||||||:
Db        373 KIFGSLAFL 381


RESULT 4
Q8LGU1
ID   Q8LGU1      PRELIMINARY;      PRT;  1294 AA.
AC   Q8LGU1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Multidrug-resistance related protein.
GN   MRP8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kolukisaoglu U.E., Bovet L., Klein M., Eggmann T., Geisler M.,
RA   Wanke D., Martinoia E., Schulz B.;
RT   "Family business: the multidrug-resistance related protein (MRP) ABC
RT   transporter genes in Arabidopsis thaliana.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ507057; CAD44995.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001140; ABC_TM_transpt.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transporter; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW   ATP-binding.
SQ   SEQUENCE   1294 AA;  143462 MW;  1638320DDB3B7C4A CRC64;

 Query Match             88.4%;  Score 38;  DB 10;  Length 1294;
 Best Local Similarity   77.8%;  Pred. No. 39;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |||||:|:|
Db        499 KVFGSIAYV 507


RESULT 5
Q9LU34
ID   Q9LU34      PRELIMINARY;      PRT;  1306 AA.
AC   Q9LU34;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Multidrug resistance-associated protein (MRP)-like, ABC-transporter-
DE   like protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=20277480; PubMed=10819329;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT   features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:131-135(2000).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AB023045; BAB01717.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001140; ABC_TM_transpt.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transporter; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW   ATP-binding; Transport.
SQ   SEQUENCE   1306 AA;  144848 MW;  73F80731E86C0D78 CRC64;

  Query Match             88.4%;  Score 38;  DB 10;  Length 1306;
  Best Local Similarity   77.8%;  Pred. No. 39;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |||||:|:|
Db        499 KVFGSIAYV 507


RESULT 6
Q9GUC0
ID   Q9GUC0      PRELIMINARY;      PRT;  333 AA.
AC   Q9GUC0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F48G7.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Clarke K., Wohldmann P., Harrison M.;
RT   "The sequence of C. elegans cosmid F48G7.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
```

```
R   EMBL; AF039044; AAG24131.1; -.
R   WormPep; F48G7.1; CE25893.
R   GO; GO:0016020; C:membrane; IEA.
R   GO; GO:0004930; F:G-protein coupled receptor activity; IEA.
R   InterPro; IPR003002; 7TM_chemrecept1.
R   InterPro; IPR000168; Nm7TM_chemrecept.
R   Pfam; PF01461; 7tm_4; 1.
W   Hypothetical protein.
Q   SEQUENCE   333 AA;  38601 MW;  77268CDC81E1A26F CRC64;

 Query Match             83.7%;  Score 36;  DB 5;  Length 333;
 Best Local Similarity   66.7%;  Pred. No. 26;
 Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 KVFGSLAFV 9
             |:||:|||:
b         13 KLFGALAFI 21


ESULT 7
7SY19
D   Q7SY19      PRELIMINARY;      PRT;   431 AA.
C   Q7SY19;
T   01-OCT-2003 (TrEMBLrel. 25, Created)
T   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
T   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E   Hypothetical protein.
S   Brachydanio rerio (Zebrafish) (Danio rerio).
C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
C   Cyprinidae; Danio.
X   NCBI_TaxID=7955;
N   [1]
P   SEQUENCE FROM N.A.
C   TISSUE=Body;
X   MEDLINE=22388257; PubMed=12477932;
A   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
A   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
A   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
A   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
A   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
A   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
A   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
A   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
A   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
A   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
A   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
A   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
A   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
A   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
A   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
A   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
A   Jones S.J., Marra M.A.;
T   "Generation and initial analysis of more than 15,000 full-length human
T   and mouse cDNA sequences.";
L   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
N   [2]
P   SEQUENCE FROM N.A.
C   TISSUE=Body;
A   Strausberg R.;
L   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
R   EMBL; BC055160; AAH55160.1; -.
W   Hypothetical protein.
Q   SEQUENCE   431 AA;  47108 MW;  8326174E46AECB1A CRC64;

 Query Match             83.7%;  Score 36;  DB 13;  Length 431;
 Best Local Similarity   66.7%;  Pred. No. 33;
 Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 KVFGSLAFV 9
             |::|||||:
b        364 KIYGSLAFL 372


RESULT 8
Q7VNU0
ID   Q7VNU0      PRELIMINARY;      PRT;   384 AA.
AC   Q7VNU0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable 3-phenylpropionic acid transporter.
GN   HCAT OR HD0391.
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000HP / ATCC 700724;
RA   Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA   Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT   "The complete genome sequence of Haemophilus ducreyi.";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017152; AAP95359.1; -.
KW   Complete proteome.
SQ   SEQUENCE   384 AA;  43328 MW;  9ACCA5A01EC48B6A CRC64;

  Query Match             81.4%;  Score 35;  DB 16;  Length 384;
  Best Local Similarity   66.7%;  Pred. No. 49;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              ::|||:|||
Db        135 RLFGSIAFV 143


RESULT 9
Q7XWZ9
ID   Q7XWZ9      PRELIMINARY;      PRT;  1708 AA.
AC   Q7XWZ9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBa0079F16.4 protein.
GN   OSJNBA0079F16.4.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL731614; CAD39831.1; -.
SQ   SEQUENCE   1708 AA;  188329 MW;  48D21907D047AC96 CRC64;

  Query Match             81.4%;  Score 35;  DB 10;  Length 1708;
  Best Local Similarity   77.8%;  Pred. No. 2.3e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              :||| ||||
Db       1021 RVFGCLAFV 1029


RESULT 10
```

```
)I5N6
)   Q9I5N6      PRELIMINARY;      PRT;  4180 AA.
:   Q9I5N6;
:   01-MAR-2001 (TrEMBLrel. 16, Created)
:   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
:   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
:   Hypothetical protein PA0690.
I   PA0690.
:   Pseudomonas aeruginosa.
:   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
:   Pseudomonadaceae; Pseudomonas.
:   NCBI_TaxID=287;
I   [1]
)   SEQUENCE FROM N.A.
:   STRAIN=ATCC 15692 / PAO1;
:   MEDLINE=20437337; PubMed=10984043;
\   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
\   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
\   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
\   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
\   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
\   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
:   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
:   opportunistic pathogen.";
,   Nature 406:959-964(2000).
:   EMBL; AE004504; AAG04079.1; -.
:   PIR; G83559; G83559.
:   GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
:   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
.   InterPro; IPR001969; Aspprotease_AS.
:   InterPro; IPR008638; Haemagg_act.
:   Pfam; PF05860; Haemagg_act; 1.
:   PROSITE; PS00141; ASP_PROTEASE; 1.
I   Hypothetical protein; Complete proteome.
)   SEQUENCE   4180 AA;  430016 MW;  EB181EA3E01BC7AC CRC64;

Query Match             81.4%;  Score 35;  DB 16;  Length 4180;
Best Local Similarity   87.5%;  Pred. No. 5.8e+02;
Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

,          2 VFGSLAFV 9
             |||||||:
)       1553 VFGSLAFM 1560


3SULT 11
31VA0
)   Q81VA0      PRELIMINARY;      PRT;    43 AA.
:   Q81VA0;
[   01-JUN-2003 (TrEMBLrel. 24, Created)
?   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
[   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
3   Hypothetical protein.
J   BA0598.
3   Bacillus anthracis (strain Ames).
:   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
(   NCBI_TaxID=198094;
J   [1]
?   SEQUENCE FROM N.A.
(   MEDLINE=22608414; PubMed=12721629;
\   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
\   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
\   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
\   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
\   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
\   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
\   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
\   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
\   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
\   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
\   Fraser C.M.;
[   "The genome sequence of Bacillus anthracis Ames and comparison to
[   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017026; AAP24617.1; -.
DR   TIGR; BA0598; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   43 AA;  4607 MW;  C952AC26FE082CAF CRC64;

  Query Match             79.1%;  Score 34;  DB 16;  Length 43;
  Best Local Similarity   75.0%;  Pred. No. 8.2;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VFGSLAFV 9
              |||:|||:
Db          5 VFGALAFI 12


RESULT 12
Q8LTR7
ID   Q8LTR7      PRELIMINARY;      PRT;    59 AA.
AC   Q8LTR7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Lactococcus lactis bacteriophage ul36.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=114416;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22066040; PubMed=12069529;
RA   Labrie S., Moineau S.;
RT   "Complete Genomic Sequence of Bacteriophage ul36: Demonstration of
RT   Phage Heterogeneity within the P335 Quasi-Species of Lactococcal
RT   Phages.";
RL   Virology 296:308-320(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Labrie S., Moineau S.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF349457; AAM75779.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   59 AA;  6386 MW;  3C7BCFA8CE1206A3 CRC64;

  Query Match             79.1%;  Score 34;  DB 9;  Length 59;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              ||||::||:
Db         29 KVFGTVAFL 37


RESULT 13
Q8LTL8
ID   Q8LTL8      PRELIMINARY;      PRT;    59 AA.
AC   Q8LTL8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
GN   ORF25.
OS   Lactococcus bacteriophage 4268.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=201847;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Trotter M., Ross R.P., Fitzgerald G., Coffey A.;
RT   "Sequence analysis of an obligately lytic P335 bacteriophage 4268.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF489521; AAM83064.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   59 AA;  6476 MW;  27617FA8CE1206A2 CRC64;

  Query Match             79.1%;  Score 34;  DB 9;  Length 59;
```

```
 Best Local Similarity   66.7%;  Pred. No. 11;
 Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 KVFGSLAFV 9
               ||||::||:
b         29 KVFGTVAFL 37


ESULT 14
83CR5
D   Q83CR5      PRELIMINARY;      PRT;   133 AA.
C   Q83CR5;
T   01-JUN-2003 (TrEMBLrel. 24, Created)
T   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E   Hypothetical protein.
N   CBU1042.
S   Coxiella burnetii.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
C   Coxiellaceae; Coxiella.
X   NCBI_TaxID=777;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=Nine Mile phase I / RSA 493;
X   MEDLINE=22608657; PubMed=12704232;
A   Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
A   Nelson W.C., Ward N.L., Tettelin H., Davidsen T.M., Beanan M.J.,
A   DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
A   Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
A   Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
T   "Complete genome sequence of the Q-fever pathogen, Coxiella
T   burnetii.";
L   Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
R   EMBL; AE016963; AAO90558.1; -.
R   TIGR; CBU1042; -.
W   Hypothetical protein; Complete proteome.
Q   SEQUENCE   133 AA;  15025 MW;  7519EC7662A96F34 CRC64;

 Query Match             79.1%;  Score 34;  DB 16;  Length 133;
 Best Local Similarity   75.0%;  Pred. No. 27;
 Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

y          2 VFGSLAFV 9
             :||||||:
b         51 IFGSLAFM 58


ESULT 15
8GLF9
D   Q8GLF9      PRELIMINARY;      PRT;   134 AA.
C   Q8GLF9;
T   01-MAR-2003 (TrEMBLrel. 23, Created)
T   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E   Immunoreactive protein Se23.5 (Fragment).
S   Streptococcus equi.
C   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
C   Streptococcus.
X   NCBI_TaxID=1336;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=CF32;
A   Qin A., Artiushin S., Timoney J.;
T   "Identification and Genomic Organization of Genes for Immunoreactive
T   Surface Exposed Proteins of Streptococcus equi.";
L   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AY137526; AAN18296.1; -.
T   NON_TER     134    134
Q   SEQUENCE   134 AA;  15366 MW;  BC3811DD1740F87F CRC64;

 Query Match             79.1%;  Score 34;  DB 2;  Length 134;
 Best Local Similarity   77.8%;  Pred. No. 27;
 Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 KVFGSLAFV 9
              |||| | ||
Db        54 KVFGPLVFV 62


Search completed: May 17, 2004, 12:56:25
Job time : 30.7419 secs
```

GenCore version 5.1.6

ɨ protein - protein search, using sw model

ɹn on: May 17, 2004, 12:51:02 ; Search time 10.1613 Seconds
(without alignments)
85.198 Million cell updates/sec

ɨtle: US-09-458-299A-4233
ɜrfect score: 43
ɜquence: 1 KVFGSLAFV 9

ɔoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

ɜarched: 283366 seqs, 96191526 residues

ɔtal number of hits satisfying chosen parameters: 283366

ɨnimum DB seq length: 0
ɜximum DB seq length: 2000000000

ɔst-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

ɜtabase : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ɜsult No.	Score	% Query Match	Length	DB	ID	Description
1	39	90.7	1254	2	I48161	p-185 precursor -
2	39	90.7	1255	1	A24571	protein-tyrosine k
3	39	90.7	1260	1	TVRTNU	protein-tyrosine k
4	36	83.7	356	2	T28747	hypothetical prote
5	35	81.4	298	2	S53849	ribosomal protein
6	35	81.4	4180	2	G83559	hypothetical prote
7	34	79.1	454	2	T27040	hypothetical prote
8	34	79.1	822	2	G84552	probable retroelem
9	33	76.7	231	2	T20547	hypothetical prote
10	33	76.7	297	2	A72100	4-hydroxybenzoate
11	33	76.7	297	2	A86524	benzoate octapheny
12	33	76.7	335	2	T20920	hypothetical prote
13	33	76.7	343	2	T33989	hypothetical prote
14	33	76.7	370	2	T48578	hypothetical prote
15	33	76.7	505	2	T26764	hypothetical prote
16	33	76.7	616	2	B84500	probable retroelem
17	32	74.4	129	2	AB1032	probable membrane
18	32	74.4	385	2	AH0353	probable transport
19	32	74.4	492	2	AD1144	di-tripeptide tran
20	32	74.4	502	2	G71055	hypothetical prote
21	32	74.4	503	2	D75104	transmembrane tran
22	32	74.4	503	2	T43969	hypothetical prote
23	32	74.4	503	2	JQ1654	SFR1 protein - hum
24	32	74.4	561	2	AH2314	lysyl-tRNA synthet
25	32	74.4	779	2	H71301	probable membrane-
26	32	74.4	845	2	H71317	probable methyl-ac
27	32	74.4	962	2	T51924	daf-18 protein - C
28	32	74.4	965	2	T32574	hypothetical prote
29	32	74.4	1152	2	H86486	protein Ty1/copia-
30	32	74.4	1334	2	E86451	probable copia-typ
31	31	72.1	149	2	T28298	hypothetical prote
32	31	72.1	152	2	T36058	hypothetical prote
33	31	72.1	157	2	S58038	probable olfactory
34	31	72.1	265	1	BWHIXB	bexB protein - Hae
35	31	72.1	272	2	AB2988	conserved hypothet
36	31	72.1	288	2	S36953	cytochrome-c oxida
37	31	72.1	288	2	S36954	cytochrome-c oxida
38	31	72.1	288	2	S36955	cytochrome-c oxida
39	31	72.1	308	2	F72704	probable iron (III
40	31	72.1	315	2	F98295	hypothetical prote
41	31	72.1	326	2	B72282	oligopeptide ABC t
42	31	72.1	352	2	B69528	conserved hypothet
43	31	72.1	354	2	T33270	hypothetical prote
44	31	72.1	405	1	SKPSXS	secretion protein
45	31	72.1	677	2	G84614	hypothetical prote

ALIGNMENTS

RESULT 1
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MUID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

Query Match 90.7%; Score 39; DB 2; Length 1254;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
|:||||:
Db 369 KIFGSLAFL 377

RESULT 2
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA

```
;Residues: 737-1031 <SEM>
;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P
cience 230, 1132-1139, 1985
;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos
;Reference number: A44188; MUID:86070181; PMID:2999974
;Accession: A44188
;Molecule type: DNA
;Residues: 740-910 <COU1>
;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
;Accession: B44188
;Molecule type: mRNA
;Residues: 1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255 <COU2>
;Cross-references: GB:M11730; NID:g183986
;King, C.R.; Kraus, M.H.; Aaronson, S.A.
cience 229, 974-976, 1985
;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
;Reference number: I59509; MUID:85272597; PMID:2992089
;Accession: I59509
;Status: translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 832-909 <REX>
;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
ol. Cell. Biol. 7, 2597-2601, 1987
;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
;Reference number: I57622; MUID:87286898; PMID:3039351
;Accession: I57622
;Status: translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-191 <TAL>
;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
;Genetics:
;Gene: GDB:ERBB2; NGL; NEU; HER-2
;Cross-references: GDB:120613; OMIM:164870
;Map position: 17q21.1-17q21.1
;Introns: 25/1; 75/3; 147/1; 883/3
;Note: the list of introns is incomplete
;Function:
;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
;Superfamily: epidermal growth factor receptor; protein kinase homology
;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
nase
;1-21/Domain: signal sequence #status predicted <SIG>
;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
;22-653/Domain: extracellular #status predicted <EXT>
;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
;654-675/Domain: transmembrane #status predicted <TMM>
;676-1255/Domain: intracellular #status predicted <INT>
;718-983/Domain: protein kinase homology <KIN>
;726-734/Region: protein kinase ATP-binding motif
;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
;753/Active site: Lys #status predicted
;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

 Query Match             90.7%;  Score 39;  DB 1;  Length 1255;
 Best Local Similarity   77.8%;  Pred. No. 7.1;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 KVFGSLAFV 9
             |:||||||:
b        369 KIFGSLAFL 377


ESULT 3
VRTNU
rotein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
;Species: Rattus norvegicus (Norway rat)
;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, S
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no mu
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphot
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

  Query Match             90.7%;  Score 39;  DB 1;  Length 1260;
  Best Local Similarity   77.8%;  Pred. No. 7.1;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 KVFGSLAFV 9
             |:||||||:
Db       373 KIFGSLAFL 381


RESULT 4
T28747
hypothetical protein F48G7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28747
R;Clarke, K.; Wohldmann, P.; Harrison, M.
submitted to the EMBL Data Library, January 1998
A;Description: The sequence of C. elegans cosmid F48G7.
A;Reference number: Z20517
A;Accession: T28747
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-356 <CLA>
A;Cross-references: EMBL:AF039044; PIDN:AAC47951.1; GSPDB:GN00023; CESP:F48G7.1
A;Experimental source: strain Bristol N2; clone F48G7
C;Genetics:
A;Gene: CESP:F48G7.1
A;Map position: 5
A;Introns: 64/3; 148/3; 220/1; 301/2

  Query Match             83.7%;  Score 36;  DB 2;  Length 356;
  Best Local Similarity   66.7%;  Pred. No. 8.4;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 KVFGSLAFV 9
             |:||:|||:
Db        13 KLFGALAFI 21


RESULT 5
S53849
```

ibosomal protein S3 - Acanthamoeba castellanii mitochondrion
;Species: mitochondrion Acanthamoeba castellanii
;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
;Accession: S53849
;Burger, G.; Plante, I.; Lonergan, K.M.; Gray, M.W.
. Mol. Biol. 245, 522-537, 1995
;Title: The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba castellanii: comp
;Reference number: S53825; MUID:95147275; PMID:7844823
;Accession: S53849
;Status: nucleic acid sequence not shown; translation not shown
:Molecule type: DNA
:Residues: 1-298 <BUR>
:Cross-references: GB:U12386; NID:g562028; PIDN:AAD11841.1; PID:g562053
:Experimental source: strain Neff; ATCC 30010
:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
;Genetics:
;Genome: mitochondrion
;Genetic code: SGC6
;Keywords: mitochondrion

Query Match 81.4%; Score 35; DB 2; Length 298;
Best Local Similarity 87.5%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KVFGSLAF 8
| ||||||
Db 250 KAFGSLAF 257

:SULT 6
:3559
'pothetical protein PA0690 [imported] - Pseudomonas aeruginosa (strain PAO1)
Species: Pseudomonas aeruginosa
Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
Accession: G83559
Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
iman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
ture 406, 959-964, 2000
Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
Reference number: A82950; MUID:20437337; PMID:10984043
Accession: G83559
Status: preliminary
Molecule type: DNA
Residues: 1-4180 <STO>
Cross-references: GB:AE004504; GB:AE004091; NID:g9946568; PIDN:AAG04079.1; GSPDB:GN001
Experimental source: strain PAO1
Genetics:
Gene: PA0690

Query Match 81.4%; Score 35; DB 2; Length 4180;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VFGSLAFV 9
|||||||:
Db 1553 VFGSLAFM 1560

:SULT 7
:7040
'pothetical protein Y49E10.9 - Caenorhabditis elegans
Species: Caenorhabditis elegans
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
Accession: T27040
Barlow, K.
ibmitted to the EMBL Data Library, August 1997
Reference number: Z20303
Accession: T27040
Status: preliminary; translated from GB/EMBL/DDBJ
Molecule type: DNA
Residues: 1-454 <WIL>
Cross-references: EMBL:Z98866; PIDN:CAB11549.1; GSPDB:GN00021; CESP:Y49E10.9
A;Experimental source: clone Y49E10
C;Genetics:
A;Gene: CESP:Y49E10.9
A;Map position: 3
A;Introns: 17/3; 125/1; 170/2; 260/3; 284/3; 302/2; 326/1; 396/1

Query Match 79.1%; Score 34; DB 2; Length 454;
Best Local Similarity 77.8%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
|:|| ||||
Db 439 KLFGLLAFV 447

RESULT 8
G84552
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84552
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84552
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-822 <STO>
A;Cross-references: GB:AE002093; NID:g4914370; PIDN:AAD32906.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g17490
A;Map position: 2

Query Match 79.1%; Score 34; DB 2; Length 822;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KVFGSLAFV 9
|||||: :|
Db 387 KVFGSICYV 395

RESULT 9
T20547
hypothetical protein F07C6.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T20547
R;Steward, C.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19290
A;Accession: T20547
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-231 <WIL>
A;Cross-references: EMBL:Z69659; PIDN:CAA93484.1; GSPDB:GN00022; CESP:F07C6.3
A;Experimental source: clone F07C6
C;Genetics:
A;Gene: CESP:F07C6.3
A;Map position: 4
A;Introns: 14/3; 47/3; 68/3; 123/3; 149/3; 179/3
C;Superfamily: Caenorhabditis elegans hypothetical protein F07C6.3

Query Match 76.7%; Score 33; DB 2; Length 231;
Best Local Similarity 75.0%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVFGSLAF 8
:||||:||
Db 44 QVFGSIAF 51

```
ESULT 10
72100
-hydroxybenzoate octaprenyltransferase CP0496 [imported] - Chlamydophila pneumoniae (st
;Alternate names: benzoate octaphenyltransferase
;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
;Accession: A72100; E81569
;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
ature Genet. 21, 385-389, 1999
;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
;Reference number: A72000; MUID:99206606; PMID:10192388
;Accession: A72100
;Status: preliminary
;Molecule type: DNA
;Residues: 1-297 <ARN>
;Cross-references: GB:AE001611; GB:AE001363; NID:g4376529; PIDN:AAD18414.1; PID:g437653
;Experimental source: strain CWL029
;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
ucleic Acids Res. 28, 1397-1406, 2000
;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
:Reference number: A81500; MUID:20150255; PMID:10684935
:Accession: E81569
:Status: preliminary
:Molecule type: DNA
:Residues: 1-297 <REA>
:Cross-references: GB:AE002211; GB:AE002161; NID:g7189412; PIDN:AAF38326.1; PID:g718941
:Experimental source: strain AR39, HL cells
:Genetics:
:Gene: ubiA; CP0496
:Superfamily: 4-hydroxybenzoate octaprenyltransferase

 Query Match             76.7%;  Score 33;  DB 2;  Length 297;
 Best Local Similarity   75.0%;  Pred. No. 30;
 Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VFGSLAFV 9
              |||::|||
Db         49 VFGAIAFV 56


ESULT 11
36524
enzoate octaphenyltransferase [imported] - Chlamydophila pneumoniae (strain J138)
;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
;Accession: A86524
;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
ucleic Acids Res. 28, 2311-2314, 2000
;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
;Reference number: A86491; MUID:20330349; PMID:10871362
;Accession: A86524
;Status: preliminary
;Molecule type: DNA
;Residues: 1-297 <STO>
;Cross-references: GB:BA000008; NID:g8978638; PIDN:BAA98475.1; GSPDB:GN00142
;Experimental source: strain J138
;Genetics:
;Gene: ubiA
;Superfamily: 4-hydroxybenzoate octaprenyltransferase

 Query Match             76.7%;  Score 33;  DB 2;  Length 297;
 Best Local Similarity   75.0%;  Pred. No. 30;
 Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VFGSLAFV 9
              |||::|||
Db         49 VFGAIAFV 56


ESULT 12
```

```
T20920
hypothetical protein F14H3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20920
R;McMurray, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19347
A;Accession: T20920
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-335 <WIL>
A;Cross-references: EMBL:Z83105; PIDN:CAB05480.1; GSPDB:GN00023; CESP:F14H3.1
A;Experimental source: clone F14H3
C;Genetics:
A;Gene: CESP:F14H3.1
A;Map position: 5
A;Introns: 83/3; 117/2; 185/3; 249/3

  Query Match             76.7%;  Score 33;  DB 2;  Length 335;
  Best Local Similarity   66.7%;  Pred. No. 34;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |||| |:|:
Db         12 KVFGVLSFI 20


RESULT 13
T33989
hypothetical protein Y40B10B.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33989
R;Harmon, G.; Courtney, L.; Langston, Y.; Drone, K.
submitted to the EMBL Data Library, February 1999
A;Description: The sequence of C. elegans cosmid Y40B10B.
A;Reference number: Z21451
A;Accession: T33989
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-343 <HAR>
A;Cross-references: EMBL:AF125961; PIDN:AAD14739.1; GSPDB:GN00023; CESP:Y40B10B.2
A;Experimental source: strain Bristol N2; clone Y40B10B
C;Genetics:
A;Gene: CESP:Y40B10B.2
A;Map position: 5
A;Introns: 64/3; 86/3; 122/2; 190/3; 224/3

  Query Match             76.7%;  Score 33;  DB 2;  Length 343;
  Best Local Similarity   66.7%;  Pred. No. 35;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:|| |:||
Db         12 KIFGILSFV 20


RESULT 14
T48578
hypothetical protein T31B5.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48578
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24490
A;Accession: T48578
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <BEV>
A;Cross-references: EMBL:AL163491
A;Experimental source: cultivar Columbia; BAC clone T31B5
```

:Genetics:
:Map position: 5
:Introns: 119/3
:Note: T31B5.130

Query Match 76.7%; Score 33; DB 2; Length 370;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 KVFGSLAF 8
|:|||| |
17 KIFGSLPF 24

:SULT 15
:6764
/pothetical protein Y39E4B.5 - Caenorhabditis elegans
·Species: Caenorhabditis elegans
·Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
·Accession: T26764
·Barlow, K.
ıbmitted to the EMBL Data Library, September 1999
·Reference number: Z20261
Accession: T26764
Status: preliminary; translated from GB/EMBL/DDBJ
Molecule type: DNA
Residues: 1-505 <WIL>
Cross-references: EMBL:AL110487; PIDN:CAB54427.1; CESP:Y39E4B.5
Experimental source: clone Y39E4B
Genetics:
Gene: CESP:Y39E4B.5
Introns: 39/2; 61/3; 212/2; 298/2; 426/3
Superfamily: glucose transport protein

Query Match 76.7%; Score 33; DB 2; Length 505;
Best Local Similarity 55.6%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

1 KVFGSLAFV 9
::|||: ||
426 RIFGSMCFV 434

:arch completed: May 17, 2004, 12:57:47
›b time : 10.1613 secs

GenCore version 5.1.6

)M protein - protein search, using sw model

Run on: May 17, 2004, 12:50:37 ; Search time 6.96774 Seconds
(without alignments)
67.257 Million cell updates/sec

Title: US-09-458-299A-4233
Perfect score: 43
Sequence: 1 KVPGSLAFV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	90.7	1254	1	ERB2_MESAU	Q60553 mesocricetu
2	39	90.7	1255	1	ERB2_HUMAN	P04626 homo sapien
3	39	90.7	1257	1	ERB2_RAT	P06494 rattus norv
4	35	81.4	298	1	RT03_ACACA	P46754 acanthamoeb
5	32	74.4	213	1	MDCG_XANCP	Q8p4u3 xanthomonas
6	32	74.4	436	1	VU10_HSV6U	Q01348 human herpe
7	32	74.4	561	1	SYK_ANASP	Q8ypw9 anabaena sp
8	31	72.1	265	1	BXB1_HAEIN	P19390 haemophilus
9	31	72.1	265	1	BXB2_HAEIN	P19391 haemophilus
10	31	72.1	265	1	BXB3_HAEIN	P22235 haemophilus
11	31	72.1	405	1	GSPF_PSEAE	Q00513 pseudomonas
12	31	72.1	502	1	VL2_HPV48	Q80925 human papil
13	31	72.1	506	1	SYK_SYNEL	Q8dma9 synechococc
14	31	72.1	960	1	FTSK_VIBCH	Q84i33 vibrio chol
15	31	72.1	1035	1	DPOL_RHCM6	O71121 rhesus cyto
16	30	69.8	172	1	DSBB_XANAC	Q8pnq6 xanthomonas
17	30	69.8	242	1	CD8A_BOVIN	P31783 bos taurus
18	30	69.8	264	1	LEP_RICPR	Q9ze32 rickettsia
19	30	69.8	265	1	CTRC_NEIMB	P32015 neisseria m
20	30	69.8	267	1	PSD_HELPY	O25911 helicobacte
21	30	69.8	298	1	NU1M_ARTSF	Q37714 artemia san
22	30	69.8	310	1	IPY2_YEAST	P28239 saccharomyc
23	30	69.8	372	1	VGLI_HSV23	P06764 herpes simp
24	30	69.8	372	1	VGLI_HSV2H	P13291 herpes simp
25	30	69.8	409	1	NU4M_CAEEL	P24892 caenorhabdi
26	30	69.8	456	1	PROY_SALTY	P37460 salmonella
27	30	69.8	457	1	PROY_ECOLI	P77327 escherichia
28	30	69.8	466	1	LEU2_VIBVU	Q8ded9 vibrio vuln
29	30	69.8	466	1	LEU2_VIBVY	Q7mp79 vibrio vuln
30	30	69.8	467	1	LEU2_VIBCH	Q9kp81 vibrio chol
31	30	69.8	469	1	LEU2_PHOLL	Q7n127 photorhabdu
32	30	69.8	471	1	LEU2_VIBPA	Q87ss9 vibrio para
33	30	69.8	476	1	LEU2_YERPE	Q8zih0 yersinia pe
34	30	69.8	482	1	G33A_DROME	Q9vka5 drosophila
35	30	69.8	485	1	LEU2_RHOBA	Q7uia7 rhodopirell
36	30	69.8	487	1	ACH6_CAEEL	Q93149 caenorhabdi
37	30	69.8	535	1	YFH6_YEAST	P43590 saccharomyc
38	30	69.8	540	1	YFE0_YEAST	P43562 saccharomyc
39	30	69.8	609	1	BAP2_YEAST	P38084 saccharomyc
40	30	69.8	670	1	KGPA_BOVIN	P00516 bos taurus
41	30	69.8	670	1	KGPA_HUMAN	Q13976 homo sapien
42	30	69.8	670	1	KGPA_RABIT	O77676 oryctolagus
43	30	69.8	678	1	CICL_RABIT	P51804 oryctolagus
44	30	69.8	686	1	KGPB_BOVIN	P21136 bos taurus
45	30	69.8	686	1	KGPB_HUMAN	P14619 homo sapien

ALIGNMENTS

```
RESULT 1
ERB2_MESAU
ID   ERB2_MESAU     STANDARD;      PRT;  1254 AA.
AC   Q60553;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN   ERBB2 OR NEU.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nerve;
RX   MEDLINE=94193007; PubMed=7908275;
RA   Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA   Yamazaki Y., Ishikawa T.;
RT   "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL   Gene 140:251-255(1994).
CC   -!- FUNCTION: Essential component of a neuregulin-receptor complex,
CC       althought neuregulins do not interact with it alone. GP30 is a
CC       potential ligand for this receptor. Not activated by EGF, TGF-
CC       alpha and amphiregulin (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: Heterodimer with each of the other ERBB receptors
CC       (Potential). Interacts with PRKCABP (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: Ligand-binding increases phosphorylation on tyrosine
CC       residues.
CC   -!- SIMILARITY: Belongs to the EGF receptor family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D16295; BAA03801.1; -.
DR   PIR; I48161; I48161.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000494; EGFR_L_domain.
DR   InterPro; IPR006211; Furin-like.
DR   InterPro; IPR006212; Furin_repeat.
DR   InterPro; IPR009030; Grow_fac_recep.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   InterPro; IPR004019; YLP_motif.
DR   Pfam; PF00757; Furin-like; 1.
```

```
)R   Pfam; PF00069; pkinase; 1.
)R   Pfam; PF01030; Recep_L_domain; 2.
)R   Pfam; PF02757; YLP; 2.
)R   PRINTS; PR00109; TYRKINASE.
)R   ProDom; PD000001; Prot_kinase; 1.
)R   SMART; SM00261; FU; 4.
)R   SMART; SM00219; TyrKc; 1.
)R   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
)R   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
)R   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
)W   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
)W   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
)W   Proto-oncogene; Disease mutation.
)T   SIGNAL        1     21       POTENTIAL.
)T   CHAIN        22   1254       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
)T   DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
)T   TRANSMEM    653    675       POTENTIAL.
)T   DOMAIN      676   1254       CYTOPLASMIC (POTENTIAL).
)T   DOMAIN      158    368       CYS-RICH.
)T   DOMAIN      472    644       CYS-RICH.
)T   DOMAIN      720    987       PROTEIN KINASE.
)T   NP_BIND     726    734       ATP (BY SIMILARITY).
)T   BINDING     753    753       ATP (BY SIMILARITY).
)T   ACT_SITE    845    845       BY SIMILARITY.
)T   DISULFID    195    204       BY SIMILARITY.
)T   DISULFID    199    212       BY SIMILARITY.
)T   DISULFID    236    244       BY SIMILARITY.
)T   DISULFID    240    252       BY SIMILARITY.
)T   DISULFID    255    264       BY SIMILARITY.
)T   DISULFID    268    295       BY SIMILARITY.
)T   DISULFID    299    311       BY SIMILARITY.
)T   DISULFID    315    331       BY SIMILARITY.
)T   DISULFID    334    338       BY SIMILARITY.
)T   DISULFID    511    520       BY SIMILARITY.
)T   DISULFID    515    528       BY SIMILARITY.
)T   DISULFID    531    540       BY SIMILARITY.
)T   DISULFID    544    560       BY SIMILARITY.
)T   DISULFID    563    576       BY SIMILARITY.
)T   DISULFID    567    584       BY SIMILARITY.
)T   DISULFID    587    596       BY SIMILARITY.
)T   DISULFID    600    623       BY SIMILARITY.
)T   DISULFID    626    634       BY SIMILARITY.
)T   DISULFID    630    642       BY SIMILARITY.
)T   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
)T   MOD_RES    1247   1247       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
)T   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
)T   VARIANT     658    658       V -> E (IN ONCOGENIC NEU).
)T   VARIANT     659    659       V -> E (IN ONCOGENIC NEU).
)Q   SEQUENCE   1254 AA;  138252 MW;  974C3791C21F2BE1 CRC64;

  Query Match             90.7%;  Score 39;  DB 1;  Length 1254;
  Best Local Similarity   77.8%;  Pred. No. 3.9;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

)Y          1 KVFGSLAFV 9
                |:||||||:
)b        369 KIFGSLAFL 377


(ESULT 2
:RB2_HUMAN
:D   ERB2_HUMAN     STANDARD;      PRT;  1255 AA.
\C   P04626;
)T   13-AUG-1987 (Rel. 05, Created)
)T   13-AUG-1987 (Rel. 05, Last sequence update)
)T   10-OCT-2003 (Rel. 42, Last annotation update)
)E   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE   surface receptor HER2) (MLN 19).
GN   ERBB2 OR HER2 OR NGL OR NEU.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86118663; PubMed=3003577;
RA   Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA   Saito T., Toyoshima K.;
RT   "Similarity of protein encoded by the human c-erb-B-2 gene to
RT   epidermal growth factor receptor.";
RL   Nature 319:230-234(1986).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT ALA-1170.
RX   MEDLINE=86070181; PubMed=2999974;
RA   Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA   McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA   Francke U., Levinson A., Ullrich A.;
RT   "Tyrosine kinase receptor with extensive homology to EGF receptor
RT   shares chromosomal location with neu oncogene.";
RL   Science 230:1132-1139(1985).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS CYS-452; VAL-655 AND ALA-1170.
RA   Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
RA   Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA   Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 737-1031 FROM N.A.
RX   MEDLINE=86016729; PubMed=2995967;
RA   Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT   "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT   c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT   human salivary gland adenocarcinoma.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN   [5]
RP   VARIANTS VAL-654 AND VAL-655.
RX   MEDLINE=93194196; PubMed=8095488;
RA   Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT   "Characterization of a new allele of the human ERBB2 gene by allele-
RT   specific competition hybridization.";
RL   Genomics 15:426-429(1993).
CC   -!- FUNCTION: Essential component of a neuregulin-receptor complex,
CC       althought neuregulins do not interact with it alone. GP30 is a
CC       potential ligand for this receptor. Not activated by EGF, TGF-
CC       alpha and amphiregulin.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: Heterodimer with each of the other ERBB receptors
CC       (Potential). Interacts with PRKCABP (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: Ligand-binding increases phosphorylation on tyrosine
CC       residues (By similarity).
CC   -!- POLYMORPHISM: There are fours alleles due to the variations in
CC       positions 654 and 655. Allele B1 (Ile-654/Ile-655) has a frequency
CC       of 0.782; allele B2 (Ile-654/Val-655) has a frequency of 0.206;
CC       allele B3 (Val-654/Val-655) has a frequency of 0.012.
CC   -!- SIMILARITY: Belongs to the EGF receptor family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11767; AAA35808.1; -.
DR   EMBL; M11761; AAA35808.1; JOINED.
DR   EMBL; M11762; AAA35808.1; JOINED.
```

```
R   EMBL; M11763; AAA35808.1; JOINED.
R   EMBL; M11764; AAA35808.1; JOINED.
R   EMBL; M11765; AAA35808.1; JOINED.
R   EMBL; M11766; AAA35808.1; JOINED.
R   EMBL; M11730; AAA75493.1; -.
R   EMBL; M12036; AAA35978.1; -.
R   EMBL; AY208911; AAO18082.1; -.
R   EMBL; X03363; CAA27060.1; -.
R   PIR; A24571; A24571.
R   PDB; 1N8Z; 18-FEB-03.
R   PDB; 1QR1; 01-JAN-00.
R   Genew; HGNC:3430; ERBB2.
R   MIM; 164870; -.
R   GO; GO:0005012; F:Neu/ErbB-2 receptor activity; TAS.
R   GO; GO:0004716; F:receptor signaling protein tyrosine kinase . . .; TAS.
R   GO; GO:0008283; P:cell proliferation; TAS.
R   GO; GO:0007167; P:enzyme linked receptor protein signaling pa. . .; TAS.
R   GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.
R   GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
R   InterPro; IPR000494; EGFR_L_domain.
R   InterPro; IPR006211; Furin-like.
R   InterPro; IPR006212; Furin_repeat.
R   InterPro; IPR009030; Grow_fac_recep.
R   InterPro; IPR000719; Prot_kinase.
R   InterPro; IPR001245; Tyr_pkinase.
R   InterPro; IPR008266; Tyr_pkinase_AS.
R   InterPro; IPR004019; YLP_motif.
R   Pfam; PF00757; Furin-like; 1.
R   Pfam; PF00069; pkinase; 1.
R   Pfam; PF01030; Recep_L_domain; 2.
R   Pfam; PF02757; YLP; 2.
R   PRINTS; PR00109; TYRKINASE.
R   ProDom; PD000001; Prot_kinase; 1.
R   SMART; SM00261; FU; 4.
R   SMART; SM00219; TyrKc; 1.
R   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
R   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
R   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
W   Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
W   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
W   Polymorphism; 3D-structure.
F   SIGNAL        1     21       POTENTIAL.
F   CHAIN         22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
F   DOMAIN        22    652       EXTRACELLULAR (POTENTIAL).
F   TRANSMEM     653    675       POTENTIAL.
F   DOMAIN       676   1255       CYTOPLASMIC (POTENTIAL).
F   DOMAIN       720    987       PROTEIN KINASE.
F   NP_BIND      726    734       ATP (BY SIMILARITY).
F   BINDING      753    753       ATP (BY SIMILARITY).
F   ACT_SITE     845    845       BY SIMILARITY.
F   DISULFID     195    204       BY SIMILARITY.
F   DISULFID     199    212       BY SIMILARITY.
F   DISULFID     220    227       BY SIMILARITY.
F   DISULFID     224    235       BY SIMILARITY.
F   DISULFID     236    244       BY SIMILARITY.
F   DISULFID     240    252       BY SIMILARITY.
F   DISULFID     255    264       BY SIMILARITY.
F   DISULFID     268    295       BY SIMILARITY.
F   DISULFID     299    311       BY SIMILARITY.
F   DISULFID     315    331       BY SIMILARITY.
F   DISULFID     334    338       BY SIMILARITY.
F   DISULFID     511    520       BY SIMILARITY.
F   DISULFID     515    528       BY SIMILARITY.
F   DISULFID     531    540       BY SIMILARITY.
F   DISULFID     544    560       BY SIMILARITY.
F   DISULFID     563    576       BY SIMILARITY.
F   DISULFID     567    584       BY SIMILARITY.
F   DISULFID     587    596       BY SIMILARITY.
F   DISULFID     600    623       BY SIMILARITY.
F   DISULFID     626    634       BY SIMILARITY.
F   DISULFID     630    642       BY SIMILARITY.
F   MOD_RES     1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
F   MOD_RES     1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT  CARBOHYD      68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD     124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD     187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD     259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD     530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD     571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD     629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  VARIANT      452    452       W -> C.
FT                                /FTId=VAR_016317.
FT  VARIANT      654    654       I -> V (in allele B3; dbSNP:1801201).
FT                                /FTId=VAR_004077.
FT  VARIANT      655    655       I -> V (in allele B2 and allele B3;
FT                                dbSNP:1801200).
FT                                /FTId=VAR_004078.
FT  VARIANT     1170   1170       P -> A.
FT                                /FTId=VAR_016318.
SQ  SEQUENCE   1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             90.7%;  Score 39;  DB 1;  Length 1255;
  Best Local Similarity   77.8%;  Pred. No. 3.9;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:||||||:
Db        369 KIFGSLAFL 377


RESULT 3
ERBB2_RAT
ID   ERBB2_RAT       STANDARD;      PRT;  1257 AA.
AC   P06494;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE   receptor-related protein).
GN   ERBB2 OR NEU.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=86118662; PubMed=3945311;
RA   Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT   "The neu oncogene encodes an epidermal growth factor receptor-related
RT   protein.";
RL   Nature 319:226-230(1986).
RN   [2]
RP   SEQUENCE OF 852-905 FROM N.A.
RC   TISSUE=Sciatic nerve;
RX   MEDLINE=91222560; PubMed=2025425;
RA   Lai C., Lemke G.;
RT   "An extended family of protein-tyrosine kinase genes differentially
RT   expressed in the vertebrate nervous system.";
RL   Neuron 6:691-704(1991).
RN   [3]
RP   STRUCTURE BY NMR OF 650-668.
RX   MEDLINE=92155181; PubMed=1346763;
RA   Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA   Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT   "Three dimensional structure of the transmembrane region of the proto-
RT   oncogenic and oncogenic forms of the neu protein.";
RL   EMBO J. 11:43-48(1992).
CC   -!- FUNCTION: Essential component of a neuregulin-receptor complex,
CC       althought neuregulins do not interact with it alone. GP30 is a
CC       potential ligand for this receptor. Not activated by EGF, TGF-
CC       alpha and amphiregulin.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: Heterodimer with each of the other ERBB receptors. The
```

```
CC       constitutively activated oncogenic variant forms a homodimer.
CC       Interacts with PRKCABP (By similarity).
CC    -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC    -!- PTM: Ligand-binding increases phosphorylation on tyrosine
CC       residues (By similarity).
CC    -!- SIMILARITY: Belongs to the EGF receptor family.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; X03362; CAA27059.1; ALT_INIT.
DR    PIR; A24562; TVRTNU.
DR    PDB; 1IIJ; 27-JUN-01.
DR    PDB; 1N8Y; 18-FEB-03.
DR    InterPro; IPR000494; EGFR_L_domain.
DR    InterPro; IPR006211; Furin-like.
DR    InterPro; IPR006212; Furin_repeat.
DR    InterPro; IPR009030; Grow_fac_recep.
DR    InterPro; IPR000719; Prot_kinase.
DR    InterPro; IPR001245; Tyr_pkinase.
DR    InterPro; IPR008266; Tyr_pkinase_AS.
DR    InterPro; IPR004019; YLP_motif.
DR    Pfam; PF00757; Furin-like; 1.
DR    Pfam; PF00069; pkinase; 1.
DR    Pfam; PF01030; Recep_L_domain; 2.
DR    Pfam; PF02757; YLP; 2.
DR    PRINTS; PR00109; TYRKINASE.
DR    ProDom; PD000001; Prot_kinase; 1.
DR    SMART; SM00261; FU; 4.
DR    SMART; SM00219; TyrKc; 1.
DR    PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR    PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR    PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW    Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW    Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW    Proto-oncogene; Disease mutation; 3D-structure.
FT    SIGNAL        1     21       POTENTIAL.
FT    CHAIN        22   1257       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT    DOMAIN       22    654       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM    655    677       POTENTIAL.
FT    DOMAIN      678   1257       CYTOPLASMIC (POTENTIAL).
FT    DOMAIN      159    369       CYS-RICH.
FT    DOMAIN      473    646       CYS-RICH.
FT    DOMAIN      722    989       PROTEIN KINASE.
FT    NP_BIND     728    736       ATP (BY SIMILARITY).
FT    BINDING     755    755       ATP (BY SIMILARITY).
FT    ACT_SITE    847    847       BY SIMILARITY.
FT    DISULFID    196    205       BY SIMILARITY.
FT    DISULFID    200    213       BY SIMILARITY.
FT    DISULFID    221    228       BY SIMILARITY.
FT    DISULFID    225    236       BY SIMILARITY.
FT    DISULFID    237    245       BY SIMILARITY.
FT    DISULFID    241    253       BY SIMILARITY.
FT    DISULFID    256    265       BY SIMILARITY.
FT    DISULFID    269    296       BY SIMILARITY.
FT    DISULFID    300    312       BY SIMILARITY.
FT    DISULFID    316    332       BY SIMILARITY.
FT    DISULFID    335    339       BY SIMILARITY.
FT    DISULFID    513    522       BY SIMILARITY.
FT    DISULFID    517    530       BY SIMILARITY.
FT    DISULFID    533    542       BY SIMILARITY.
FT    DISULFID    546    562       BY SIMILARITY.
FT    DISULFID    565    578       BY SIMILARITY.
FT    DISULFID    569    586       BY SIMILARITY.
FT    DISULFID    589    598       BY SIMILARITY.
FT    DISULFID    602    625       BY SIMILARITY.
FT    DISULFID    628    636       BY SIMILARITY.
FT    DISULFID    632    644       BY SIMILARITY.
FT    MOD_RES    1141   1141       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT    MOD_RES    1250   1250       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT    CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    VARIANT     661    661       V -> E (IN ONCOGENIC NEU).
SQ    SEQUENCE   1257 AA;  138831 MW;  6129264583011402 CRC64;

  Query Match             90.7%;  Score 39;  DB 1;  Length 1257;
  Best Local Similarity   77.8%;  Pred. No. 3.9;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |:||||||:
Db        370 KIFGSLAFL 378


RESULT 4
RT03_ACACA
ID    RT03_ACACA     STANDARD;      PRT;   298 AA.
AC    P46754;
DT    01-NOV-1995 (Rel. 32, Created)
DT    01-NOV-1995 (Rel. 32, Last sequence update)
DT    30-MAY-2000 (Rel. 39, Last annotation update)
DE    Mitochondrial ribosomal protein S3.
GN    RPS3.
OS    Acanthamoeba castellanii (Amoeba).
OG    Mitochondrion.
OC    Eukaryota; Acanthamoebidae; Acanthamoeba.
OX    NCBI_TaxID=5755;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=ATCC 30010 / Neff;
RX    MEDLINE=95147275; PubMed=7844823;
RA    Burger G., Plante I., Lonergan K.M., Gray M.W.;
RT    "The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba
RT    castellanii: complete sequence, gene content and genome
RT    organization.";
RL    J. Mol. Biol. 245:522-537(1995).
CC    -!- SUBCELLULAR LOCATION: Mitochondrial.
CC    -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; U12386; AAD11841.1; -.
DR    PIR; S53849; S53849.
DR    InterPro; IPR009019; KH_prok.
DR    InterPro; IPR001351; Ribosomal_S3_C.
DR    InterPro; IPR008282; Ribosomal_S3_N.
DR    Pfam; PF00189; Ribosomal_S3_C; 1.
DR    Pfam; PF00417; Ribosomal_S3_N; 1.
DR    PROSITE; PS00548; RIBOSOMAL_S3; FALSE_NEG.
KW    Ribosomal protein; Mitochondrion.
SQ    SEQUENCE   298 AA;  36060 MW;  29415935EE187DE6 CRC64;

  Query Match             81.4%;  Score 35;  DB 1;  Length 298;
  Best Local Similarity   87.5%;  Pred. No. 6.8;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAF 8
              | ||||||
Db        250 KAFGSLAF 257
```

```
ESULT 5
MDCG_XANCP
ID   MDCG_XANCP      STANDARD;      PRT;   213 AA.
AC   Q8P4U3;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoribosyl-dephospho-CoA transferase (EC 2.7.7.-) (Holo-ACP
DE   synthase).
GN   MDCG OR XCC3613.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: Transfers 2-(5''-triphosphoribosyl)-3'-
CC       dephosphocoenzyme-A to the apo-acyl carrier protein of the
CC       malonate decarboxylase to yield holo-acyl carrier protein (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: 2'-(5''-triphosphoribosyl)-3'-dephospho-CoA +
CC       apo-ACP = holo-ACP + diphosphate.
CC   -!- SIMILARITY: Belongs to the mdcG family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE012482; AAM42883.1; -.
DR   HAMAP; MF_00650; -; 1.
KW   Transferase; Nucleotidyltransferase; Complete proteome.
FT   ACT_SITE    135    135       BY SIMILARITY.
FT   ACT_SITE    137    137       BY SIMILARITY.
SQ   SEQUENCE   213 AA;  22901 MW;  C784E3CD7B061AEF CRC64;

 Query Match            74.4%;  Score 32;  DB 1;  Length 213;
 Best Local Similarity   75.0%;  Pred. No. 21;
 Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAF 8
              :|||| ||
Db        114 RVFGSFAF 121


ESULT 6
U10_HSV6U
ID   VU10_HSV6U      STANDARD;      PRT;   436 AA.
AC   Q01348;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   U10 protein.
GN   U10 OR SFR1 OR P1RF0.
OS   Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Roseolovirus.
OX   NCBI_TaxID=10370;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95266321; PubMed=7747482;
RA   Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA   Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;
RT   "The DNA sequence of human herpesvirus-6: structure, coding content,
RT   and genome evolution.";
RL   Virology 209:29-51(1995).
RN   [2]
RP   SEQUENCE OF 1-407 FROM N.A.
RX   MEDLINE=92333249; PubMed=1321206;
RA   Efstathiou S., Lawrence G.L., Brown C.M., Barrell B.G.;
RT   "Identification of homologues to the human cytomegalovirus US22 gene
RT   family in human herpesvirus 6.";
RL   J. Gen. Virol. 73:1661-1671(1992).
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-6 AND
CC       HSV-7 U10 AND HCMV UL31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83413; CAA58437.1; -.
DR   EMBL; D10082; BAA20953.1; -.
DR   InterPro; IPR007578; DUF570.
DR   Pfam; PF04489; DUF570; 1.
SQ   SEQUENCE   436 AA;  50122 MW;  6ADCA71AC9ECE87E CRC64;

  Query Match            74.4%;  Score 32;  DB 1;  Length 436;
  Best Local Similarity   77.8%;  Pred. No. 41;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              ||||:|| |
Db        283 KVFGTLATV 291


RESULT 7
SYK_ANASP
ID   SYK_ANASP       STANDARD;      PRT;   561 AA.
AC   Q8YPW9;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).
GN   LYSS OR ALL4071.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
CC   -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
```

```
'C       + L-lysyl-tRNA(Lys).
'C    -!- COFACTOR: Binds 3 magnesium ions per subunit (By similarity).
'C    -!- SUBUNIT: Homodimer (By similarity).
'C    -!- SUBCELLULAR LOCATION: Cytoplasmic.
'C    -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
'C    ----------------------------------------------------------------------
'C    This SWISS-PROT entry is copyright. It is produced through a collaboration
'C    between the Swiss Institute of Bioinformatics and the EMBL outstation -
'C    the European Bioinformatics Institute. There are no restrictions on its
'C    use by non-profit institutions as long as its content is in no way
'C    modified and this statement is not removed. Usage by and for commercial
'C    entities requires a license agreement (See http://www.isb-sib.ch/announce/
'C    or send an email to license@isb-sib.ch).
'C    ----------------------------------------------------------------------
'R    EMBL; AP003595; BAB75770.1; -.
'R    PIR; AH2314; AH2314.
'R    HAMAP; MF_00252; -; 1.
'R    InterPro; IPR008994; Nucleic_acid_OB.
'R    InterPro; IPR004364; tRNA-synt_2.
'R    InterPro; IPR002312; tRNA-synt_asp.
'R    InterPro; IPR002313; tRNA-synt_lys_2.
'R    InterPro; IPR004365; tRNA_anti.
'R    InterPro; IPR006195; tRNA_ligase_II.
'R    Pfam; PF00152; tRNA-synt_2; 1.
'R    Pfam; PF01336; tRNA_anti; 1.
'R    PRINTS; PR01042; TRNASYNTHASP.
'R    PRINTS; PR00982; TRNASYNTHLYS.
'R    TIGRFAMs; TIGR00499; lysS_bact; 1.
'R    PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
'W    Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
'W    Metal-binding; Magnesium; Complete proteome.
'T    METAL       409    409       MAGNESIUM 1 (BY SIMILARITY).
'T    METAL       416    416       MAGNESIUM 1 AND 2 (BY SIMILARITY).
'Q    SEQUENCE   561 AA;  63676 MW;  92FF80E13632EAA3 CRC64;

 Query Match             74.4%;  Score 32;  DB 1;  Length 561;
 Best Local Similarity   75.0%;  Pred. No. 52;
 Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

)y          1 KVFGSLAF 8
              :||| |||
)b         65 RVFGKLAF 72


!ESULT 8
3XB1_HAEIN
!D   BXB1_HAEIN     STANDARD;      PRT;   265 AA.
\C   P19390;
)T   01-NOV-1990 (Rel. 16, Created)
)T   01-NOV-1990 (Rel. 16, Last sequence update)
)T   15-MAR-2004 (Rel. 43, Last annotation update)
)E   Capsule polysaccharide export inner-membrane protein bexB.
3N   BEXB.
)S   Haemophilus influenzae.
)C   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
)C   Pasteurellaceae; Haemophilus.
)X   NCBI_TaxID=727;
3N   [1]
3P   SEQUENCE FROM N.A.
3C   STRAIN=RM 153 / Serotype B;
3X   MEDLINE=90170850; PubMed=2137816;
3A   Kroll J.S., Moxon E.R.;
3T   "Capsulation in distantly related strains of Haemophilus influenzae
3T   type b: genetic drift and gene transfer at the capsulation locus.";
3L   J. Bacteriol. 172:1374-1379(1990).
)C   -!- FUNCTION: May form an ATP-driven capsule polysaccharide export
)C       apparatus, in association with the bexA, bexC and bexD proteins.
)C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
)C       (Potential).
)C   -!- SIMILARITY: Belongs to the ABC-2 integral membrane protein family.
)C   ----------------------------------------------------------------------
)C   This SWISS-PROT entry is copyright. It is produced through a collaboration
)C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M33787; AAA24945.1; -.
DR   InterPro; IPR000412; ABC_transpt2.
DR   Pfam; PF01061; ABC2_membrane; 1.
DR   PRINTS; PR00164; ABC2TRNSPORT.
DR   PROSITE; PS00890; ABC2_MEMBRANE; 1.
KW   Transport; Polysaccharide transport; Bacterial capsule;
KW   Inner membrane; Transmembrane.
FT   TRANSMEM     37     57       POTENTIAL.
FT   TRANSMEM     64     84       POTENTIAL.
FT   TRANSMEM    118    138       POTENTIAL.
FT   TRANSMEM    151    171       POTENTIAL.
FT   TRANSMEM    178    198       POTENTIAL.
FT   TRANSMEM    235    255       POTENTIAL.
SQ   SEQUENCE   265 AA;  30181 MW;  0A436FF824CD25C1 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 265;
  Best Local Similarity   55.6%;  Pred. No. 42;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              |::|:|:||
Db        180 KIWGTLSFV 188


RESULT 9
BXB2_HAEIN
ID   BXB2_HAEIN     STANDARD;      PRT;   265 AA.
AC   P19391;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Capsule polysaccharide export inner-membrane protein bexB.
GN   BEXB.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RM 926 / Serotype B;
RX   MEDLINE=90170850; PubMed=2137816;
RA   Kroll J.S., Moxon E.R.;
RT   "Capsulation in distantly related strains of Haemophilus influenzae
RT   type b: genetic drift and gene transfer at the capsulation locus.";
RL   J. Bacteriol. 172:1374-1379(1990).
CC   -!- FUNCTION: May form an ATP-driven capsule polysaccharide export
CC       apparatus, in association with the bexA, bexC and bexD proteins.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: Belongs to the ABC-2 integral membrane protein family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M33788; AAA24946.1; -.
DR   InterPro; IPR000412; ABC_transpt2.
DR   Pfam; PF01061; ABC2_membrane; 1.
DR   PRINTS; PR00164; ABC2TRNSPORT.
DR   PROSITE; PS00890; ABC2_MEMBRANE; 1.
KW   Transport; Polysaccharide transport; Bacterial capsule;
KW   Inner membrane; Transmembrane.
FT   TRANSMEM     37     57       POTENTIAL.
```

```
T   TRANSMEM     64     84       POTENTIAL.
T   TRANSMEM    118    138       POTENTIAL.
T   TRANSMEM    151    171       POTENTIAL.
T   TRANSMEM    178    198       POTENTIAL.
T   TRANSMEM    235    255       POTENTIAL.
Q   SEQUENCE   265 AA;  30108 MW;  FBA0C49386E91274 CRC64;

 Query Match             72.1%;  Score 31;  DB 1;  Length 265;
 Best Local Similarity   55.6%;  Pred. No. 42;
 Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

          1 KVFGSLAFV 9
            |::|:|:||
        180 KIWGTLSFV 188


ESULT 10
(B3_HAEIN
D   BXB3_HAEIN     STANDARD;      PRT;   265 AA.
C   P22235;
T   01-AUG-1991 (Rel. 19, Created)
T   01-AUG-1991 (Rel. 19, Last sequence update)
T   15-MAR-2004 (Rel. 43, Last annotation update)
E   Capsule polysaccharide export inner-membrane protein bexB.
N   BEXB.
S   Haemophilus influenzae.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
C   Pasteurellaceae; Haemophilus.
X   NCBI_TaxID=727;
R   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=Eagan / Serotype B;
X   MEDLINE=91186821; PubMed=2082145;
A   Kroll J.S., Loynds B., Brophy L.N., Moxon E.R.;
T   "The bex locus in encapsulated Haemophilus influenzae: a chromosomal
T   region involved in capsule polysaccharide export.";
L   Mol. Microbiol. 4:1853-1862(1990).
C   -!- FUNCTION: May form an ATP-driven capsule polysaccharide export
C       apparatus, in association with the bexA, bexC and bexD proteins.
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
C       (Potential).
C   -!- SIMILARITY: Belongs to the ABC-2 integral membrane protein family.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; X54987; CAA38733.1; -.
R   PIR; S12234; BWHIXB.
R   InterPro; IPR000412; ABC_transpt2.
R   Pfam; PF01061; ABC2_membrane; 1.
R   PRINTS; PR00164; ABC2TRNSPORT.
R   PROSITE; PS00890; ABC2_MEMBRANE; 1.
K   Transport; Polysaccharide transport; Bacterial capsule;
K   Inner membrane; Transmembrane.
T   TRANSMEM     37     57       POTENTIAL.
T   TRANSMEM     64     84       POTENTIAL.
T   TRANSMEM    118    138       POTENTIAL.
T   TRANSMEM    151    171       POTENTIAL.
T   TRANSMEM    178    198       POTENTIAL.
T   TRANSMEM    235    255       POTENTIAL.
Q   SEQUENCE   265 AA;  30195 MW;  E3A4F181C4B3203E CRC64;

 Query Match             72.1%;  Score 31;  DB 1;  Length 265;
 Best Local Similarity   55.6%;  Pred. No. 42;
 Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 KVFGSLAFV 9
             |::|:|:||
```

```
Db         180 KIWGTLSFV 188


RESULT 11
GSPF_PSEAE
ID   GSPF_PSEAE     STANDARD;      PRT;   405 AA.
AC   Q00513;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   General secretion pathway protein F.
GN   XCPS OR PA3102.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=92269572; PubMed=1588814;
RA   Bally M., Filloux A., Akrim M., Ball G., Lazdunski A., Tommassen J.;
RT   "Protein secretion in Pseudomonas aeruginosa: characterization of
RT   seven xcp genes and processing of secretory apparatus components by
RT   prepilin peptidase.";
RL   Mol. Microbiol. 6:1121-1131(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC       EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF A VARIETY
CC       OF ENZYMES ACROSS THE OUTER MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE PULF/OUTF/EXEF/XPSF/XCPS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62666; CAA44534.1; -.
DR   EMBL; AE004734; AAG06490.1; -.
DR   PIR; S25385; SKPSXS.
DR   InterPro; IPR003004; Bac_GSPF.
DR   InterPro; IPR001992; Bact_secr_systII.
DR   Pfam; PF00482; GSPII_F; 1.
DR   PRINTS; PR00812; BCTERIALGSPF.
DR   PROSITE; PS00874; T2SP_F; 1.
KW   Transport; Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM    169    189       POTENTIAL.
FT   TRANSMEM    219    239       POTENTIAL.
FT   TRANSMEM    377    397       POTENTIAL.
SQ   SEQUENCE   405 AA;  44061 MW;  BEE59B58724C167E CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 405;
  Best Local Similarity   66.7%;  Pred. No. 62;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         1 KVFGSLAFV 9
             | :|:||||
```

```
)b        222 KAWGALAFV 230


:ESULT 12
'L2_HPV48
D   VL2_HPV48      STANDARD;      PRT;   502 AA.
:C   Q80925;
:T   15-JUL-1998 (Rel. 36, Created)
:T   15-JUL-1998 (Rel. 36, Last sequence update)
:T   15-JUL-1998 (Rel. 36, Last annotation update)
E   Minor capsid protein L2.
N   L2.
S   Human papillomavirus type 48.
C   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
C   Papillomavirus.
X   NCBI_TaxID=40538;
N   [1]
P   SEQUENCE FROM N.A.
A   Delius H.;
L   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
'C   ---------------------------------------------------------------------
'C   This SWISS-PROT entry is copyright. It is produced through a collaboration
'C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
'C   the European Bioinformatics Institute. There are no restrictions on its
'C   use by non-profit institutions as long as its content is in no way
'C   modified and this statement is not removed. Usage by and for commercial
'C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; U31789; AAA79469.1; -.
R   InterPro; IPR000784; Late_L2.
R   Pfam; PF00513; late_protein_L2; 1.
W   Coat protein; Late protein.
Q   SEQUENCE   502 AA;  54435 MW;  C42DE45A7E73EE29 CRC64;

 Query Match             72.1%;  Score 31;  DB 1;  Length 502;
 Best Local Similarity   62.5%;  Pred. No. 76;
 Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

)y          1 KVFGSLAF 8
              |:|||| :
)b         43 KIFGSLVY 50


:ESULT 13
:YK_SYNEL
:D   SYK_SYNEL      STANDARD;      PRT;   506 AA.
:C   Q8DMA9;
)T   10-OCT-2003 (Rel. 42, Created)
)T   10-OCT-2003 (Rel. 42, Last sequence update)
)T   10-OCT-2003 (Rel. 42, Last annotation update)
)E   Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).
:N   LYSS OR TLL0212.
)S   Synechococcus elongatus (Thermosynechococcus elongatus).
)C   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
)X   NCBI_TaxID=32046;
:N   [1]
:P   SEQUENCE FROM N.A.
:C   STRAIN=BP-1;
:X   MEDLINE=22225144; PubMed=12240834;
:A   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
:A   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
:A   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
:A   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
:T   "Complete genome structure of the thermophilic cyanobacterium
:T   Thermosynechococcus elongatus BP-1.";
:L   DNA Res. 9:123-130(2002).
:C   -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
:C       + L-lysyl-tRNA(Lys).
:C   -!- COFACTOR: Binds 3 magnesium ions per subunit (By similarity).
:C   -!- SUBUNIT: Homodimer (By similarity).
:C   -!- SUBCELLULAR LOCATION: Cytoplasmic.
:C   -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005369; BAC07765.1; -.
DR   HAMAP; MF_00252; -; 1.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR004364; tRNA-synt_2.
DR   InterPro; IPR002313; tRNA-synt_lys_2.
DR   InterPro; IPR004365; tRNA_anti.
DR   InterPro; IPR006195; tRNA_ligase_II.
DR   Pfam; PF00152; tRNA-synt_2; 1.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   PRINTS; PR00982; TRNASYNTHLYS.
DR   TIGRFAMs; TIGR00499; lysS_bact; 1.
DR   PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Metal-binding; Magnesium; Complete proteome.
FT   METAL       411    411       MAGNESIUM 1 (BY SIMILARITY).
FT   METAL       418    418       MAGNESIUM 1 AND 2 (BY SIMILARITY).
SQ   SEQUENCE   506 AA;  56353 MW;  837861ED74C8F1F5 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 506;
  Best Local Similarity   62.5%;  Pred. No. 76;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAF 8
              ::|| |||
Db         70 RIFGKLAF 77


RESULT 14
FTSK_VIBCH
ID   FTSK_VIBCH     STANDARD;      PRT;   960 AA.
AC   Q84I33; Q9KQU5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA translocase ftsK.
GN   FTSK OR VC1903.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
RN   [2]
RP   SEQUENCE OF 446-917 FROM N.A.
RX   MEDLINE=22450551; PubMed=12562793;
RA   Herz K., Vimont S., Padan E., Berche P.;
RT   "Roles of NhaA, NhaB, and NhaD Na(+)/H(+) antiporters in survival of
RT   Vibrio cholerae in a saline environment.";
RL   J. Bacteriol. 185:1236-1244(2003).
CC   -!- FUNCTION: DNA motor protein, which is both required to move DNA
CC       out of the region of the septum during cell division and for the
CC       septum formation. Tracks DNA in an ATP-dependent manner by
```

```
C       generating positive supercoils in front of it and negative
C       supercoils behind it (By similarity).
C   -!- SUBUNIT: Homohexamer. This suggests the formation of a ring
C       between the two cells at the septum that surrounds DNA (By
C       similarity).
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Located at the
C       septum. The large C-terminal part of the protein is cytoplasmic
C       (Potential).
C   -!- SIMILARITY: Contains 1 FtsK domain.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; AE004266; AAF95051.1; -.
R   EMBL; AF489522; AAO37927.1; -.
R   PIR; A82142; A82142.
R   TIGR; VC1903; -.
R   HAMAP; MF_01809; -; 1.
R   InterPro; IPR002543; FtsK_SpoIIIE.
R   Pfam; PF01580; FtsK_SpoIIIE; 1.
R   PROSITE; PS50901; FTSK; 1.
W   Chromosome partition; Cell division; ATP-binding; DNA-binding;
W   Transmembrane; Complete proteome.
T   TRANSMEM     33     55       POTENTIAL.
T   TRANSMEM     84    106       POTENTIAL.
T   TRANSMEM    119    141       POTENTIAL.
T   TRANSMEM    146    168       POTENTIAL.
T   TRANSMEM    173    195       POTENTIAL.
T   DOMAIN      601    814       FTSK.
T   NP_BIND     618    625       ATP (POTENTIAL).
Q   SEQUENCE   960 AA;  105887 MW;  0AA778438B7D8970 CRC64;

 Query Match             72.1%;  Score 31;  DB 1;  Length 960;
 Best Local Similarity   85.7%;  Pred. No. 1.4e+02;
 Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

y          2 VFGSLAF 8
             ||||||:
b         86 VFGSLAY 92


ESULT 15
POL_RHCM6
D   DPOL_RHCM6     STANDARD;      PRT;  1035 AA.
C   O71121;
T   15-DEC-1998 (Rel. 37, Created)
T   15-DEC-1998 (Rel. 37, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   DNA polymerase (EC 2.7.7.7).
N   UL54.
S   Rhesus cytomegalovirus (strain 68-1) (RhCMV).
C   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C   Betaherpesvirinae; Cytomegalovirus.
X   NCBI_TaxID=103930;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=98118459; PubMed=9454707;
A   Swanson R., Bergquam E., Wong S.W.;
T   "Characterization of rhesus cytomegalovirus genes associated with
T   anti-viral susceptibility.";
L   Virology 240:338-348(1998).
C   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
C       + {DNA}(N).
C   -!- SUBCELLULAR LOCATION: Nuclear.
C   -!- SIMILARITY: Belongs to the DNA polymerase type-B family.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; AF033184; AAC05256.1; -.
DR  InterPro; IPR006172; DNA_pol_B.
DR  InterPro; IPR006134; DNA_pol_B_dom.
DR  InterPro; IPR006133; DNA_pol_B_exo.
DR  Pfam; PF00136; DNA_pol_B; 1.
DR  Pfam; PF03104; DNA_pol_B_exo; 1.
DR  PRINTS; PR00106; DNAPOLB.
DR  SMART; SM00486; POLBc; 1.
DR  PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW  Transferase; DNA-directed DNA polymerase; DNA replication;
KW  DNA-binding; Nuclear protein.
SQ  SEQUENCE   1035 AA;  116595 MW;  4E320D2F062D90C1 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 1035;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 KVFGSLAFV 9
              ||||||  :
Db        799 KVFGSLMMI 807


Search completed: May 17, 2004, 12:57:00
Job time : 7.96774 secs
```

GenCore version 5.1.6

4 protein - protein search, using sw model

ın on: May 17, 2004, 12:46:47 ; Search time 41.5161 Seconds
(without alignments)
61.252 Million cell updates/sec

.tle: US-09-458-299A-4239
erfect score: 42
equence: 1 VVLGVVFGV 9

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	42	100.0	9	4	AAG88792	Aag88792 HER2/neu
2	42	100.0	9	4	AAB75857	Aab75857 Tumour as
3	41	97.6	9	2	AAW70053	Aaw70053 HER-2/neu
4	41	97.6	9	2	AAY47712	Aay47712 Immunogen
5	41	97.6	9	4	AAB99702	Aab99702 HLA A2 bi
6	41	97.6	9	4	AAG88791	Aag88791 HER2/neu
7	41	97.6	9	4	AAG89000	Aag89000 HER2/neu
8	41	97.6	9	4	AAG88769	Aag88769 HER2/neu
9	41	97.6	9	4	AAG88467	Aag88467 HER2/NEU
10	41	97.6	9	5	ABG79080	Abg79080 Human HER
11	41	97.6	9	7	ADA49641	Ada49641 Multi-epi
12	41	97.6	15	3	AAY98853	Aay98853 HLA class
13	41	97.6	15	4	AAG88468	Aag88468 HER2/NEU
14	41	97.6	15	4	AAG88418	Aag88418 HER2/NEU
15	41	97.6	15	4	AAG89023	Aag89023 Her2/neu
16	41	97.6	15	4	AAG88636	Aag88636 HER2/NEU
17	41	97.6	22	2	AAW53134	Aaw53134 HER-2/neu
18	41	97.6	44	6	ABU10039	Abu10039 Proto-onc
19	41	97.6	50	7	ABR82072	Abr82072 Proximal
20	41	97.6	69	2	AAW53132	Aaw53132 HER-2/neu
21	41	97.6	144	7	ADA49445	Ada49445 Multi-epi
22	41	97.6	147	7	ADA49447	Ada49447 Multi-epi
23	41	97.6	148	7	ADA49443	Ada49443 Multi-epi
24	41	97.6	151	6	ABU10036	Abu10036 Toxic sho
25	41	97.6	238	6	ABU10035	Abu10035 Toxic sho
26	41	97.6	275	7	ABR82071	Abr82071 Human Her
27	41	97.6	1200	3	AAB21208	Aab21208 Human HER
28	41	97.6	1223	5	AAU98923	Aau98923 Human bre
29	41	97.6	1253	7	ADC35106	Adc35106 Human bre
30	41	97.6	1255	2	AAW01111	Aaw01111 HER-2/neu
31	41	97.6	1255	2	AAW92406	Aaw92406 Human HER
32	41	97.6	1255	3	AAY92620	Aay92620 Human her
33	41	97.6	1255	3	AAY84780	Aay84780 Amino aci
34	41	97.6	1255	3	AAB21198	Aab21198 Human HER
35	41	97.6	1255	4	AAB60167	Aab60167 HER2 tran
36	41	97.6	1255	4	AAG88267	Aag88267 HER2/neu
37	41	97.6	1255	4	AAE12130	Aae12130 Human tyr
38	41	97.6	1255	4	AAB85458	Aab85458 Human HER
39	41	97.6	1255	5	AAE20479	Aae20479 Human Her
40	41	97.6	1255	5	AAU77114	Aau77114 Human Her
41	41	97.6	1255	5	AAE26349	Aae26349 Human HER
42	41	97.6	1255	5	AAE26366	Aae26366 Human Her
43	41	97.6	1255	5	AAU74545	Aau74545 Human HER
44	41	97.6	1255	5	AAM51143	Aam51143 Human Her
45	41	97.6	1255	5	AAE24067	Aae24067 Human Her

ALIGNMENTS

RESULT 1
AAG88792
ID AAG88792 standard; peptide; 9 AA.
XX
AC AAG88792;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu A2 supermotif crossbinding peptide #36.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Example 2; Page 180; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the

prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample form a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour-tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in the exemplification of the present invention

Sequence 9 AA;

```
Query Match             100.0%;  Score 42;  DB 4;  Length 9;
Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              |||||||||
Db          1 VVLGVVFGV 9
```

RESULT 2
AAB75857
AAB75857 standard; peptide; 9 AA.

AAB75857;

10-APR-2001 (first entry)

Tumour associated antigen Her2/neu HLA-A2 binding peptide.

Human leukocyte antigen; HLA; major histocompatibility complex; MHC; cytotoxic T lymphocyte; CTL; human class I MHC; immunogenic; HLA binding peptide; immune response; glycoprotein; cytostatic; virucide; hepatotropic; antiinflammatory; anti-HIV; vaccine; human immunodeficiency virus; protozoacide; viral infection; cancer; prostate cancer; hepatitis B; hepatitis C; human papilloma virus; HPV; cytomegalovirus; CMV; acquired immunodeficiency syndrome; AIDS; renal carcinoma; cervical carcinoma; lymphoma; malaria; condyloma acuminatum.

Homo sapiens.

WO200100225-A1.

04-JAN-2001.

28-JUN-2000; 2000WO-US017842.

29-JUN-1999; 99US-0141422P.

(EPIM-) EPIMMUNE INC.

Sette A, Sidney J, Southwood S;

WPI; 2001-112389/12.

Composition comprising human leukocyte antigen binding peptide which comprises isolated, prepared epitope useful for treating viral infections such as acquired immunodeficiency syndrome, and cancer.

Claim 1; Page 42; 58pp; English.

The present invention describes a composition (I) which comprises at least one human leukocyte antigen (HLA) binding peptide comprising an isolated, prepared epitope comprising one of 547 8-11 residue amino acid sequences (S1), given in AAB75803 to AAB76349. (I) has cytostatic, virucide, hepatotropic, antiinflammatory, anti-HIV (human immunodeficiency virus) and protozoacide activities, which can be used in vaccine production and is an inducer of cytotoxic T-cell response. (I) is useful for inducing a cytotoxic T cell response against a preselected antigen in a patient expressing a specific major histocompatibility complex (MHC) class I allele, by contacting cytotoxic T cells (CTLs) from the patient with (I). (I) is useful as a vaccine to treat and/or prevent viral infection and cancer such as prostate cancer, hepatitis B, hepatitis C, human papilloma virus (HPV) infection, cytomegalovirus (CMV), acquired immunodeficiency syndrome (AIDS), renal carcinoma, cervical carcinoma, lymphoma, malaria, and condyloma acuminatum

Sequence 9 AA;

```
  Query Match             100.0%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              |||||||||
Db          1 VVLGVVFGV 9
```

RESULT 3
AAW70053
ID AAW70053 standard; peptide; 9 AA.
XX
AC AAW70053;
XX
DT 22-OCT-1998 (first entry)
XX
DE HER-2/neu derived HLA-A2.1 binding peptide 1 (residues 565-573).
XX
KW Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC; human leukocyte antigen; HLA; tumour associated antigen; cancer; antigen presenting cell; APC; immunogenic peptide; immune disorder; viral infection; AIDS; hepatitis; bacterial infection; malaria; fungal infection; tuberculosis; melanoma; HER-2/neu; cerB-2.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9833888-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US001959.
XX
PR 31-JAN-1997; 97US-0036696P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Tsai V, Southwood S, Sidney J, Sette A, Celis E;
XX
DR WPI; 1998-437445/37.
XX
PT Production of antigen-specific cytotoxic T cells - by incubating
PT immunogenic peptide(s) from antigen that binds class I major
PT histocompatibility complex molecules with pre-treated antigen presenting
PT cells.
XX
PS Example 7; Page 77; 104pp; English.
XX
CC Sequences shown in AAW70053 to AAW70075 represent peptides derived from
CC HER-2/neu (cerB-2) antigen. The peptides can bind to a human leukocyte
CC antigen (HLA), HLA-A2.1 and are used to exemplify the method of invention
CC of producing antigen-specific cytotoxic T cells (CTLs) in vitro. The
CC method comprises contacting immunogenic peptides from an antigen that
CC binds class I major histocompatibility complex (MHC) molecules with
CC antigen presenting cells (APCs) pretreated with pretreatment growth
CC factors, and incubating the APCs with purified CD8 cells in the presence
CC of at least 2 incubation growth factors, thereby producing antigen-
CC specific CTLs. A method for specifically killing target cells in a human

CC patient is also provided which comprises obtaining a fluid sample
CC containing CTLs from a patient, contacting the cytotoxic T cells with
CC APCs pretreated with pre-treatment growth factors, where the APCs
CC comprise class I MHC molecules. The pretreated APCs are incubated with
CC the cytotoxic growth factors, thereby producing activated CTLs which are
CC contacted with a carrier to form a composition. The composition can then
CC be administered to the patient. The activated CTLs can be used for
CC treating cancers, immune disorders, viral infections, AIDS, hepatitis,
CC bacterial infection, fungal infection, malaria or tuberculosis
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 4
AAY47712
ID AAY47712 standard; peptide; 9 AA.
XX
AC AAY47712;
XX
DT 01-DEC-1999 (first entry)
XX
DE Immunogenic peptide having a human leukocyte antigen binding motif #2323.
XX
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW immune response; T cell activation; major histocompatibility complex;
KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW vaccine; immunisation.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9945954-A1.
XX
PD 16-SEP-1999.
XX
PF 13-MAR-1998; 98WO-US005039.
XX
PR 13-MAR-1998; 98WO-US005039.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;
XX
DR WPI; 1999-551214/46.
XX
PT New immunogenic peptides with HLA binding motif, useful in treatment and
PT diagnosis of cancers and viral diseases.
XX
PS Claim 1; Page 120; 150pp; English.
XX
CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
CC having a human major histocompatibility complex (MHC) Class I (also known
CC as human leukocyte antigen (HLA)) binding motif. The immunogenic peptides
CC can bind to a specific HLA allele (i.e. HLA-A subtypes HLA-A2.1, A1, A3.2
CC or A24.1 or HLA-B or C) and induce a cytotoxic T cell response against
CC the antigen from which the peptide is derived. Cytotoxic T lymphocytes
CC (CTLs) which destroy antigen-bearing cells are normally induced by an
CC antigen in the form of a peptide fragment bound to a HLA molecule, rather
CC than the intact foreign antigen itself, and are particularly important in
CC tumour rejection and in fighting viral infections. The peptides are
CC therefore useful therapeutically to treat or prevent viral infections and
CC cancers in mammals (especially humans) e.g. prostate cancer, hepatitis B
CC and C, AIDS, and renal carcinoma. They can be administered as vaccines to
CC elicit an immune response in individuals susceptible or otherwise at risk
CC of viral infection or cancer, or used to treat chronic or acute
CC conditions. They are also useful diagnostically, and can be used to
CC induce a cytotoxic T cell response, by contacting a cytotoxic T cell with
CC the peptide e.g. to produce CTLs ex vivo for infusion back into a
CC patient. The polynucleotides encoding the immunogenic peptides are also
CC useful therapeutically and for immunisation as above
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 5
AAB99702
ID AAB99702 standard; peptide; 9 AA.
XX
AC AAB99702;
XX
DT 06-SEP-2001 (first entry)
XX
DE HLA A2 binding CTL epitope peptide from Her2/neu SEQ ID NO:23.
XX
KW Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL;
KW cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu;
KW MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator;
KW immunotherapy; immune response.
XX
OS Homo sapiens.
XX
PN WO200141741-A1.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-US034318.
XX
PR 13-DEC-1999; 99US-0170448P.
PR 05-APR-2000; 2000US-00543608.
PR 30-MAY-2000; 2000US-00583200.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Celis E, Keogh E;
PI Chesnut R;
XX
DR WPI; 2001-381489/40.
XX
PT Compositions for use in a vaccine for treating, e.g., breast, lung and
PT colon cancer comprises at least one peptide that comprises an isolated
PT epitope of a tumor-associated antigen.
XX
PS Claim 1; Page 76; 86pp; English.
XX
CC The present invention describes a composition (I) comprising at least one
CC peptide that comprises an isolated, prepared epitope consisting of a
CC sequence selected from 25 short amino acid sequences given in AAB99680 to
CC AAB99704. Also described are: (1) a composition (II) comprising one or
CC more peptides, and further comprising at least two epitopes selected from
CC the 25 short amino acid sequences (as above), where each of the one or
CC more peptides comprise less than 50 contiguous amino acids that have 100%
CC identity with a native peptide sequence; and (2) a vaccine composition
CC (III) comprising an epitope selected from the 25 short amino acid
CC sequences (as above) and a pharmaceutical excipient. (I) has cytostatic
CC and immunomodulatory activities and can be used in vaccine production and
CC immunotherapy. The peptide epitope compositions (I)-(II) are useful for
CC monitoring an immune response to a tumour associated antigen or when one
CC or more peptides are combined to create a vaccine (III) that stimulates
CC the cellular arm of the immune system. In particular, the vaccine

CC mediates immune responses against tumours in individuals who bear an
CC allele of the human leukocyte antigen (HLA)-A2 supertype and improve the
CC standard of care for patients being treated for breast, colon, or lung
CC cancer
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 6
AAG88791
ID AAG88791 standard; peptide; 9 AA.
XX
AC AAG88791;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu A2 supermotif crossbinding peptide #35.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Example 2; Page 180; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 7
AAG89000
ID AAG89000 standard; peptide; 9 AA.
XX
AC AAG89000;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu epitope HLA-A2 supermotif-bearing peptide #13.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Claim 1; Page 189; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by

CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 8
AAG88769
ID AAG88769 standard; peptide; 9 AA.
XX
AC AAG88769;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu A2 supermotif crossbinding peptide #13.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Example 2; Page 179; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 9
AAG88467
ID AAG88467 standard; peptide; 9 AA.
XX
AC AAG88467;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/NEU DR supermotif binding peptide core sequence #90.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Disclosure; Page 168; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and

CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 10
ABG79080
ID ABG79080 standard; peptide; 9 AA.
XX
AC ABG79080;
XX
DT 15-NOV-2002 (first entry)
XX
DE Human HER-2 class I HLA widely expressed antigen peptide #4.
XX
KW Cell penetrating peptide; cancer; tumour; melanoma; thymoma; antigen;
KW lymphoma; sarcoma; lung cancer; non-Hodgkin's lymphoma; leukaemia;
KW Hodgkin's lymphoma; uterine cancer; cervical cancer; bladder cancer;
KW kidney cancer; adenocarcinoma; breast cancer; prostate cancer;
KW ovarian cancer; pancreatic cancer; epitope; vaccine; dendritic cell;
KW tumour infiltrating lymphocyte; TIL; human leukocyte antigen; HLA;
KW cytostatic; human.
XX
OS Homo sapiens.
XX
PN WO200264057-A2.
XX
PD 22-AUG-2002.
XX
PF 15-FEB-2002; 2002WO-US005212.
XX
PR 15-FEB-2001; 2001US-0268687P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Wang R;
XX
DR WPI; 2002-627577/67.
XX
PT Novel composition for treating a disease in an animal, comprises an
PT immune effector cell and cell penetrating peptide associated with an
PT antigen or antibody.
XX
PS Disclosure; Page 18; 61pp; English.
XX
CC The invention relates to a composition (I) comprising an immune effector
CC cell and a cell penetrating peptide (CPP) associated with an antigen or
CC antibody. Also included are (1) a vaccine comprising (I), CPP associated
CC with an antigen, and a pharmaceutically acceptable carrier and (2)
CC preparing a composition for a disease, by providing (I) and CPP
CC associated with an antigen for disease, and introducing the antigen-
CC associated CPP to (I), where antigen enters into the cell. The antigens
CC are, for example, tumour antigen derived epitopes recognised by tumour
CC infiltrating lymphocytes (TIL) of HLA (human leukocyte antigen) class I
CC or II. The composition is useful for enhancing immunity in an animal to a
CC disease, by administering a mature dendritic cell comprising CPP
CC associated with an antigen to disease, to the animal, such that following
CC the administration, animal is protected from disease, where the animal
CC comprises both CD4+ and CD8+ T cells. It is also useful for treating a
CC disease (e.g. cancer, tumour, melanoma, thymoma, lymphoma, sarcoma, lung
CC cancer, non-Hodgkin's lymphoma, leukaemia, Hodgkin's lymphoma, uterine
CC cancer, cervical cancer, bladder cancer, kidney cancer, adenocarcinoma,
CC breast cancer, prostate cancer, ovarian cancer and pancreatic cancer).
CC The animal is further subjected to a cancer treatment including surgery,
CC radiation, chemotherapy or gene therapy. The administration of (I),
CC preferably dendritic cell is prior to, subsequent to or concurrent with,
CC the cancer treatment. The present sequence is a tumour antigen derived
CC epitope for inclusion in the composition of the invention
XX
SQ Sequence 9 AA;

Query Match 97.6%; Score 41; DB 5; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.4e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 1 VVLGVVFGI 9

RESULT 11
ADA49641
ID ADA49641 standard; peptide; 9 AA.
XX
AC ADA49641;
XX
DT 20-NOV-2003 (first entry)
XX
DE Multi-epitope construct specific epitope #183.
XX
KW multi-epitope; immunogenic; epitope; major histocompatibility complex;
KW MHC class I; MHC class II; junctional epitope.
XX
OS Unidentified.
XX
PN US2002119127-A1.
XX
PD 29-AUG-2002.
XX
PF 27-JUN-2001; 2001US-00894018.
XX
PR 28-DEC-1999; 99US-0173390P.
PR 28-DEC-2000; 2000WO-US035568.
PR 16-APR-2001; 2001US-0284221P.
XX
PA (SETT/) SETTE A.
PA (CHES/) CHESNUT R.
PA (LIVI/) LIVINGSTON B D.
PA (BAKE/) BAKER D M.
PA (NEWM/) NEWMAN M J.
PA (BROW/) BROWN D H.
XX
PI Sette A, Chesnut R; Livingston BD, Baker DM, Newman MJ, Brown DH;
XX
DR WPI; 2003-615704/58.
XX
PT Designing multi-epitope construct having major histocompatability complex
PT class I and II epitope nucleic acids, by selecting mixture of amino acid
PT insertions at junctions of construct to minimize junctional epitopes.
XX
PS Disclosure; Fig 19E; 78pp; English.
XX
CC The invention relates to a method of designing multi-epitope constructs
CC comprising major histocompatibility complex (MHC) class I and II (CTL)
CC epitope nucleic acids (CEN), involves sorting CEN, introducing flanking
CC amino acid residue selected from specified amino acid residues given in
CC specification at C+1 position of CEN, introducing amino acid spacer

```
C   residues between two CEN, and selecting the constructs having less
C   junctional epitopes. The method is useful for designing a multi-epitope
C   construct having multiple epitope nucleic acid. The method avoids or
C   minimises the occurrence of junctional epitopes and maximises the
C   immunogenicity and/or antigenicity of multi-epitope vaccines. The present
C   sequence represents the amino acid sequence of an epitope present in a
C   multi-epitope construct.
X
Q   Sequence 9 AA;

 Query Match             97.6%;  Score 41;  DB 7;  Length 9;
 Best Local Similarity   88.9%;  Pred. No. 1.4e+06;
 Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 VVLGVVFGV 9
             ||||||||:
b          1 VVLGVVFGI 9


ESULT 12
AY98853
D   AAY98853 standard; peptide; 15 AA.
X
C   AAY98853;
X
T   07-AUG-2000  (first entry)
X
E   HLA class II binding antigen epitope peptide #42.
X
W   Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
W   immune response; chronic viral disease; cancer; autoimmune disease;
W   rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
W   allograft rejection; allergy; lyme disease; hepatitis; prostate cancer;
W   glomerulonephritis; food hypersensitivity; malaria.
X
S   Unidentified.
X
N   WO9961916-A1.
X
D   02-DEC-1999.
X
F   28-MAY-1999;   99WO-US012066.
X
R   29-MAY-1998;   98US-0087192P.
X
A   (EPIM-) EPIMMUNE INC.
X
I   Sette A,  Southwood S,  Sidney J;
X
R   WPI; 2000-097143/08.
X
T   New compositions containing immunogenic peptide epitopes for various HLA
T   class II DR molecules useful for inducing helper T cell response.
X
S   Claim 1; Page 40; 60pp; English.
X
C   The present invention relates to a new pharmaceutical composition
C   comprising a unit dose form of a peptide, or analogue, comprising an
C   epitope selected from those represented by peptides AAY98812-Y99339 which
C   are derived from various antigens for various human leucocyte antigen
C   class DR molecules, representative of the world wide population. The
C   peptide/analogue binds to an HLA class II molecule at an IC-50 of less
C   than or equal to 1,000 nM. The pharmaceutical can be used to induce a
C   helper T cell response. The pharmaceutical focuses the immune response
C   towards selected determinants and could therefore be used in cases of
C   chronic viral diseases and cancer. Examples of diseases that can be
C   treated using the peptide containing pharmaceutical include autoimmune
C   diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
C   gravis), allograft rejection, allergies, lyme disease, hepatitis, post-
C   streptococcal endocarditis or glomerulonephritis and food
C   hypersensitivities. The peptide epitopes can be used to enhance immune
C   responses against other immunogens administered with the peptides.
C   Diseases which can be treated using immunogenic mixtures include prostate
CC  cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma, cervical
CC  carcinoma, lymphoma, and condyloma acuminatum. The peptides may also be
CC  used to make monoclonal antibodies useful as potential diagnostic or
CC  therapeutic agents. The peptides may also be useful as diagnostic
CC  reagents, for example, to determine the susceptibility of an individual
CC  to a treatment regimen. Also, the peptides may be used to predict which
CC  individuals will be at substantial risk of developing chronic infection.
CC  The selection of appropriate T and B cell epitopes should allow the
CC  development of epitope based vaccines particularly towards conserved
CC  epitopes of pathogens which are characterized by high sequence
CC  variability such as HIV, HCV and Malaria
XX
SQ  Sequence 15 AA;

  Query Match             97.6%;  Score 41;  DB 3;  Length 15;
  Best Local Similarity   88.9%;  Pred. No. 0.38;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 VVLGVVFGV 9
             ||||||||:
Db         4 VVLGVVFGI 12


RESULT 13
AAG88468
ID  AAG88468 standard; peptide; 15 AA.
XX
AC  AAG88468;
XX
DT  11-SEP-2001  (first entry)
XX
DE  HER2/NEU DR supermotif binding peptide exemplary sequence #90.
XX
KW  Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW  immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW  tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS  Homo sapiens.
OS  Synthetic.
XX
PN  WO200141787-A1.
XX
PD  14-JUN-2001.
XX
PF  11-DEC-2000; 2000WO-US033591.
XX
PR  10-DEC-1999;   99US-00458299.
XX
PA  (EPIM-) EPIMMUNE INC.
XX
PI  Fikes J,  Sette A,  Sidney J,  Southwood S,  Chesnut R,  Celis E;
PI  Keogh E;
XX
DR  WPI; 2001-374995/39.
XX
PT  An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT  cellular immune responses for the prevention and treatment of cancer.
XX
PS  Disclosure; Page 168; 199pp; English.
XX
CC  The present invention describes isolated prepared HER2/neu epitopes (I).
CC  Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC  culture in vitro and binds to a complex of an epitope (I), bound to a
CC  human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC  and a second epitope and the peptide is less than 50 contiguous amino
CC  acids that have 100% identity with a native peptide sequence of HER2/neu;
CC  (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC  excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC  ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC  immunostimulant activities, and can be used in vaccines. (I), (II) and
CC  (III) are useful for inducing cellular immune responses for the
CC  prevention and treatment of cancer. (I) and (II) are useful for
CC  monitoring or evaluating an immune response to a tumour-associated
```

CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 15 AA;

Query Match 97.6%; Score 41; DB 4; Length 15;
Best Local Similarity 88.9%; Pred. No. 0.38;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 4 VVLGVVFGI 12

RESULT 14
AAG88418
ID AAG88418 standard; peptide; 15 AA.
XX
AC AAG88418;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/NEU DR supermotif binding peptide exemplary sequence #65.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Disclosure; Page 168; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
SQ Sequence 15 AA;

Query Match 97.6%; Score 41; DB 4; Length 15;
Best Local Similarity 88.9%; Pred. No. 0.38;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||||||:
Db 5 VVLGVVFGI 13

RESULT 15
AAG89023
ID AAG89023 standard; peptide; 15 AA.
XX
AC AAG89023;
XX
DT 11-SEP-2001 (first entry)
XX
DE Her2/neu DR supertype primary binding peptide #17.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033591.
XX
PR 10-DEC-1999; 99US-00458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
PS Example 5; Page 190; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC ; and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and

immunostimulant activities, and can be used in vaccines. (I), (II) and (III) are useful for inducing cellular immune responses for the prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample form a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour-tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in the exemplification of the present invention

Sequence 15 AA;

Query Match 97.6%; Score 41; DB 4; Length 15;
Best Local Similarity 88.9%; Pred. No. 0.38;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
          1 VVLGVVFGV 9
            ||||||||:
          4 VVLGVVFGI 12
```

Search completed: May 17, 2004, 12:54:34
Job time : 43.5161 secs

GenCore version 5.1.6

M protein - protein search, using sw model

un on: May 17, 2004, 12:51:02 ; Search time 10.1613 Seconds
(without alignments)
85.198 Million cell updates/sec

itle: US-09-458-299A-4239
erfect score: 42
equence: 1 VVLGVVFGV 9

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	41	97.6	1255	1	A24571	protein-tyrosine k
2	41	97.6	2824	2	T22759	hypothetical prote
3	39	92.9	556	2	S51892	probable membrane
4	37	88.1	377	2	A69277	Na+/H+ antiporter
5	36	85.7	57	2	B81350	small hydrophobic
6	36	85.7	93	2	G64521	hypothetical prote
7	36	85.7	93	2	C71984	hypothetical prote
8	36	85.7	135	2	S49200	cytochrome b5 - co
9	36	85.7	385	2	G82789	permease XF0589 [i
10	36	85.7	429	2	A97241	permease [imported
11	36	85.7	451	2	S74833	hypothetical prote
12	36	85.7	822	2	AG1911	hypothetical prote
13	35	83.3	35	2	H84362	hypothetical prote
14	35	83.3	80	2	C86630	prophage ps1 prote
15	35	83.3	139	2	S46306	cytochrome b5 - co
16	35	83.3	149	2	T04603	hypothetical prote
17	35	83.3	159	2	A85330	hypothetical prote
18	35	83.3	204	2	AE3302	multiple antibioti
19	35	83.3	219	2	B81296	probable integral
20	35	83.3	234	2	T04604	hypothetical prote
21	35	83.3	321	2	E65097	hypothetical 35.8
22	35	83.3	321	2	B91125	probable transport
23	35	83.3	321	2	A85970	probable transport
24	35	83.3	322	2	A97657	ferrichrome ABC tr
25	35	83.3	322	2	AI2880	ABC transporter, m
26	35	83.3	399	2	D71728	proton/sodium-glut
27	35	83.3	399	2	G97727	proton/sodium-glut
28	35	83.3	437	2	S56529	probable phosphotr
29	35	83.3	437	2	AF0667	probable phosphotr
30	35	83.3	454	2	F75580	probable sugar tra
31	35	83.3	473	2	C64739	yadQ protein - Esc
32	35	83.3	473	2	G90648	probable channel t
33	35	83.3	473	2	G85499	probable channel t
34	35	83.3	523	2	S55677	glutamate transpor
35	35	83.3	524	2	S28902	glutamate transpor
36	34	81.0	166	2	G86167	protein F21B7.33 [
37	34	81.0	167	2	T00888	hypothetical prote
38	34	81.0	170	2	C90079	conserved hypothet
39	34	81.0	205	2	E83293	hypothetical prote
40	34	81.0	217	2	G95873	probable amino aci
41	34	81.0	274	2	B75518	bacitracin resista
42	34	81.0	287	2	C90023	hypothetical prote
43	34	81.0	295	1	QRECMA	chemotaxis protein
44	34	81.0	295	2	H90953	proton conductor c
45	34	81.0	295	2	AB0203	chemotaxis MotA pr

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517,'RALL',522,'S',524-654,'V',656-1169,'A',1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
;Genetics:
;Gene: GDB:ERBB2; NGL; NEU; HER-2
;Cross-references: GDB:120613; OMIM:164870
;Map position: 17q21.1-17q21.1
;Introns: 25/1; 75/3; 147/1; 883/3
;Note: the list of introns is incomplete
;Function:
;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
;Superfamily: epidermal growth factor receptor; protein kinase homology
;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
nase
;1-21/Domain: signal sequence #status predicted <SIG>
;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
;22-653/Domain: extracellular #status predicted <EXT>
;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
;654-675/Domain: transmembrane #status predicted <TMM>
;676-1255/Domain: intracellular #status predicted <INT>
;718-983/Domain: protein kinase homology <KIN>
;726-734/Region: protein kinase ATP-binding motif
;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
;753/Active site: Lys #status predicted
;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.6%; Score 41; DB 1; Length 1255;
Best Local Similarity 88.9%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

y 1 VVLGVVFGV 9
||||||||:
b 665 VVLGVVFGI 673

ESULT 2
22759
ypothetical protein F55H12.3 - Caenorhabditis elegans
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
;Accession: T22759
;Dobson, R.
ubmitted to the EMBL Data Library, October 1996
.;Reference number: Z19610
.;Accession: T22759
.;Status: preliminary; translated from GB/EMBL/DDBJ
.;Molecule type: DNA
.;Residues: 1-2824 <WIL>
.;Cross-references: EMBL:Z81091; PIDN:CAB03143.1; GSPDB:GN00019; CESP:F55H12.3
.;Experimental source: clone F55H12
';Genetics:
.;Gene: CESP:F55H12.3
.;Map position: 1
.;Introns: 87/1; 98/1; 126/2; 201/3; 343/3; 406/1; 576/3; 656/1; 825/3; 869/1; 909/1; 96
/1; 1755/2; 1800/1; 1850/3; 1896/1; 2003/3; 2035/3; 2082/3; 2119/1; 2144/1; 2200/2; 227
';Superfamily: LDL receptor ligand-binding repeat homology
';243-279/Domain: LDL receptor ligand-binding repeat homology <LDL>

Query Match 97.6%; Score 41; DB 2; Length 2824;
Best Local Similarity 88.9%; Pred. No. 28;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

y 1 VVLGVVFGV 9
:||||||||
b 2651 IVLGVVFGV 2659

ESULT 3
51892
robable membrane protein YOL105c - yeast (Saccharomyces cerevisiae)
;Alternate names: hypothetical protein HRE556; hypothetical protein O0759
';Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 03-Aug-1995 #text_change 21-Jul-2000
C;Accession: S51892; S59168; S66801
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV including
and a Delta.
A;Reference number: S51848
A;Accession: S51892
A;Molecule type: DNA
A;Residues: 1-556 <VAN>
A;Cross-references: EMBL:Z48149; NID:g663234; PID:g663247
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
Yeast 11, 1069-1075, 1995
A;Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including the
a delta element.
A;Reference number: S59156; MUID:96076631; PMID:7502582
A;Accession: S59168
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-556 <VAW>
A;Cross-references: EMBL:Z48149; NID:g663234; PIDN:CAA88155.1; PID:g663247
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
R;Durand, P.; Hilger, F.; Portetelle, D.; Vandenbol, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66791
A;Accession: S66801
A;Molecule type: DNA
A;Residues: 1-556 <DUR>
A;Cross-references: EMBL:Z74847; NID:g1419966; PID:e252294; PID:g1419967; MIPS:YOL105c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:WSC3
A;Cross-references: SGD:S0005465; MIPS:YOL105c
A;Map position: 15L
C;Keywords: transmembrane protein
F;20-36/Domain: transmembrane #status predicted <TM1>
F;317-333/Domain: transmembrane #status predicted <TM2>
F;385-401/Domain: transmembrane #status predicted <TM3>

Query Match 92.9%; Score 39; DB 2; Length 556;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
:|:||||||
Db 385 IVIGVVFGV 393

RESULT 4
A69277
Na+/H+ antiporter (napA-1) homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Jun-2000
C;Accession: A69277
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.,
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: A69277
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-377 <KLE>
A;Cross-references: GB:AE001091; GB:AE000782; NID:g2689414; PIDN:AAB91016.1; PID:g2650426
C;Superfamily: Aquifex aeolicus Na+/H+-exchanging protein napA1

Query Match 88.1%; Score 37; DB 2; Length 377;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Y          1 VVLGVVFGV 9
             :|||::|||
D         36 IVLGMIFGV 44


ESULT 5
31350
mall hydrophobic protein Cj0786 [imported] - Campylobacter jejuni (strain NCTC 11168)
;Species: Campylobacter jejuni
;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
;Accession: B81350
;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
ature 403, 665-668, 2000
;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
;Reference number: A81250; MUID:20150912; PMID:10688204
;Accession: B81350
;Status: preliminary
;Molecule type: DNA
;Residues: 1-57 <PAR>
;Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB73051.1; PID:g696823
;Experimental source: serotype O2, strain NCTC 11168
;Genetics:
;Gene: Cj0786

 Query Match             85.7%;  Score 36;  DB 2;  Length 57;
 Best Local Similarity   44.4%;  Pred. No. 6.5;
 Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Y          1 VVLGVVFGV 9
             :::||:||:
D          9 IIIGVIFGI 17


ESULT 6
4521
pothetical protein HP0015 - Helicobacter pylori (strain 26695)
Species: Helicobacter pylori
Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
Accession: G64521
Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
eterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
n, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
ature 388, 539-547, 1997
Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
Reference number: A64520; MUID:97394467; PMID:9252185
Accession: G64521
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-93 <TOM>
Cross-references: GB:AE000524; GB:AE000511; NID:g2313090; PIDN:AAD07091.1; PID:g231309

 Query Match             85.7%;  Score 36;  DB 2;  Length 93;
 Best Local Similarity   75.0%;  Pred. No. 10;
 Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Y          2 VLGVVFGV 9
             :||||||:
D         69 ILGVVFGI 76


ESULT 7
71984
pothetical protein jhp0013 - Helicobacter pylori (strain J99)
Species: Helicobacter pylori
Variety: strain J99
Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
Accession: C71984
Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
ture 397, 176-180, 1999
Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: C71984
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-93 <ARN>
A;Cross-references: GB:AE001441; GB:AE001439; NID:g4154511; PIDN:AAD05595.1; PID:g415451
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0013

  Query Match             85.7%;  Score 36;  DB 2;  Length 93;
  Best Local Similarity   75.0%;  Pred. No. 10;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VLGVVFGV 9
              :||||||:
Db         69 ILGVVFGI 76


RESULT 8
S49200
cytochrome b5 - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 05-May-2000
C;Accession: S49200; S25523
R;Napier, J.A.; Smith, M.A.; Shewry, P.R.; Stobart, A.K.
submitted to the EMBL Data Library, July 1994
A;Reference number: S49200
A;Accession: S49200
A;Molecule type: mRNA
A;Residues: 1-135 <NAP>
A;Cross-references: EMBL:X80008; NID:g510538; PIDN:CAA56318.1; PID:g510539
R;Smith, M.A.; Stobart, A.K.; Shewry, P.R.; Napier, J.A.
submitted to the EMBL Data Library, August 1992
A;Description: Isolation and expression pattern of a tobaco cytochrome b5 cDNA.
A;Reference number: S25523
A;Accession: S25523
A;Molecule type: mRNA
A;Residues: 'EF',12-14,'Q',16-24,'V',26-31,'D',33-37,'D',39-43,'D',45-61,'V',63-82,'A',8
A;Cross-references: EMBL:X68140; NID:g19852; PIDN:CAA48240.1; PID:g19853
C;Superfamily: cytochrome b5; cytochrome b5 core homology
C;Keywords: heme; iron; metalloprotein
F;5-80/Domain: cytochrome b5 core homology <CB5>
F;40,64/Binding site: heme iron (His) (axial ligands) #status predicted

  Query Match             85.7%;  Score 36;  DB 2;  Length 135;
  Best Local Similarity   66.7%;  Pred. No. 14;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::||| |||
Db        119 IILGVAFGV 127


RESULT 9
G82789
permease XF0589 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82789
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-385 <SIM>
A;Cross-references: GB:AE003904; GB:AE003849; NID:g9105433; PIDN:AAF83399.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
```

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
T.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
Thado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
I.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
C;Gene: XF0589

Query Match 85.7%; Score 36; DB 2; Length 385;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
|::||||||
Db 284 VLVGVVFGV 292

RESULT 10
A97241
Permease [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: A97241
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-429 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80716.1; PID:g15025810; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
C;Gene: CAC2772
C;Superfamily: conserved hypothetical protein HI0125

Query Match 85.7%; Score 36; DB 2; Length 429;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VLGVVFGV 9
|||::|||
Db 203 VLGIIFGV 210

RESULT 11
S74833
hypothetical protein sll0855 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
C;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S74833
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
NA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74833
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-451 <KAN>
A;Cross-references: EMBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAA17794.1; PID:g165287
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: hypothetical protein sll0855

Query Match 85.7%; Score 36; DB 2; Length 451;
Best Local Similarity 55.6%; Pred. No. 43;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
::||::|||
Db 251 IILGILFGV 259

RESULT 12
AG1911
hypothetical protein alr0841 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AG1911
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anal
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG1911
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-822 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB72798.1; PID:g17130186; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr0841

Query Match 85.7%; Score 36; DB 2; Length 822;
Best Local Similarity 77.8%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
| :||||||
Db 222 VAIGVVFGV 230

RESULT 13
H84362
hypothetical protein Vng2129h [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84362
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lia
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: H84362
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-35 <STO>
A;Cross-references: GB:AE004437; NID:g10581543; PIDN:AAG20268.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG2129H

Query Match 83.3%; Score 35; DB 2; Length 35;
Best Local Similarity 77.8%; Pred. No. 6.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VVLGVVFGV 9
||||| :||
Db 19 VVLGVTYGV 27

RESULT 14
C86630

```
rophage ps1 protein 19 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
;Species: Lactococcus lactis subsp. lactis
;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
;Accession: C86630
;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
enome Res. 11, 731-753, 2001
;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
;Reference number: A86625; MUID:21235186; PMID:11337471
;Accession: C86630
;Status: preliminary
;Molecule type: DNA
;Residues: 1-80 <STO>
;Cross-references: GB:AE005176; PID:g12722883; PIDN:AAK04141.1; GSPDB:GN00146
;Experimental source: strain IL1403
;Genetics:
;Gene: ps119

 Query Match             83.3%;  Score 35;  DB 2;  Length 80;
 Best Local Similarity   62.5%;  Pred. No. 14;
 Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

y          2 VLGVVFGV 9
             :|||:||:
b         58 ILGVIFGI 65


ESULT 15
46306
ytochrome b5 - common tobacco
;Species: Nicotiana tabacum (common tobacco)
;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 05-May-2000
;Accession: S46306; S33157
;Smith, M.A.; Stobart, A.K.; Shewry, P.R.; Napier, J.A.
lant Mol. Biol. 25, 527-537, 1994
;Title: Tobacco cytochrome b(5): cDNA isolation, expression analysis and in vitro prote
;Reference number: S46306; MUID:94325476; PMID:8049375
;Accession: S46306
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-139 <SMI>
;Cross-references: EMBL:X71441; NID:g296385; PIDN:CAA50575.1; PID:g296386
;Superfamily: cytochrome b5; cytochrome b5 core homology
;Keywords: heme; iron; metalloprotein
;8-83/Domain: cytochrome b5 core homology <CB5>
;43,67/Binding site: heme iron (His) (axial ligands) #status predicted

 Query Match             83.3%;  Score 35;  DB 2;  Length 139;
 Best Local Similarity   55.6%;  Pred. No. 22;
 Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

y          1 VVLGVVFGV 9
             ::||| ||:
o        122 IILGVAFGI 130


earch completed: May 17, 2004, 12:57:48
ob time : 11.1613 secs
```

GenCore version 5.1.6

M protein - protein search, using sw model

un on: May 17, 2004, 12:50:37 ; Search time 6.96774 Seconds
(without alignments)
67.257 Million cell updates/sec

itle: US-09-458-299A-4239
erfect score: 42
equence: 1 VVLGVVFGV 9

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 141681 seqs, 52070155 residues

otal number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	41	97.6	1255	1	ERB2_HUMAN	P04626 homo sapien
2	39	92.9	556	1	WSC3_YEAST	Q12215 saccharomyc
3	37	88.1	162	1	CAV2_FUGRU	Q9ygm9 fugu rubrip
4	36	85.7	135	1	CYS5_TOBAC	P49099 nicotiana t
5	36	85.7	137	1	CT24_HUMAN	Q9buv8 homo sapien
6	35	83.3	136	1	CYB5_TOBAC	P49098 nicotiana t
7	35	83.3	275	1	COBS_COREF	Q8fnq1 corynebacte
8	35	83.3	320	1	ALX_SHIFL	Q83q35 shigella fl
9	35	83.3	321	1	ALX_ECO57	Q8xaj0 escherichia
10	35	83.3	321	1	ALX_ECOL6	Q8fde1 escherichia
11	35	83.3	321	1	ALX_ECOLI	P42601 escherichia
12	35	83.3	437	1	SGCC_ECOLI	P39365 escherichia
13	35	83.3	473	1	CLCA_ECO57	P58244 escherichia
14	35	83.3	473	1	CLCA_ECOL6	Q8f115 escherichia
15	35	83.3	473	1	CLCA_ECOLI	P37019 escherichia
16	35	83.3	473	1	CLCA_SHIFL	P59639 shigella fl
17	35	83.3	523	1	EAA3_MOUSE	P51906 mus musculu
18	35	83.3	524	1	EAA3_BOVIN	Q95135 bos taurus
19	35	83.3	524	1	EAA3_RABIT	P31597 oryctolagus
20	34	81.0	274	1	UPK_DEIRA	Q9rx61 deinococcus
21	34	81.0	295	1	MOTA_ECOLI	P09348 escherichia
22	34	81.0	322	1	ALX_SALTI	Q8z315 salmonella
23	34	81.0	322	1	ALX_SALTY	Q8zlx2 salmonella
24	34	81.0	430	1	G43B_DROME	Q9v4q0 drosophila
25	34	81.0	463	1	YDJN_ECOLI	P77529 escherichia
26	34	81.0	482	1	YGFU_ECOLI	Q46821 escherichia
27	34	81.0	523	1	EAA3_RAT	P51907 rattus norv
28	34	81.0	533	1	YE91_HELPJ	Q9zjc8 helicobacte
29	34	81.0	533	1	YE91_HELPY	O26024 helicobacte
30	34	81.0	655	1	NDC1_YEAST	P32500 saccharomyc
31	34	81.0	1742	1	GUNA_CALSA	P22534 caldocellum
32	33	78.6	40	1	YSXC_SULAC	P39477 sulfolobus
33	33	78.6	108	1	Y869_ARCFU	O29392 archaeoglob
34	33	78.6	129	1	CT24_MOUSE	Q9cqt9 mus musculu
35	33	78.6	271	1	ELO3_MOUSE	O35949 mus musculu
36	33	78.6	367	1	Y797_METJA	Q58207 methanococc
37	33	78.6	589	1	Y015_MYCGE	P47261 mycoplasma
38	33	78.6	815	1	NAH1_HUMAN	P19634 homo sapien
39	33	78.6	917	1	VGLB_HSVB2	P12641 bovine herp
40	33	78.6	942	1	DPO1_CHLAU	O08307 chloroflexu
41	32	76.2	87	1	Y476_MYCLE	Q9cb43 mycobacteri
42	32	76.2	118	1	Y914_AQUAE	O67059 aquifex aeo
43	32	76.2	229	1	Y790_METJA	Q58200 methanococc
44	32	76.2	295	1	MOTA_SALTY	P55891 salmonella
45	32	76.2	311	1	OPPB_BACSU	P24138 bacillus su

ALIGNMENTS

RESULT 1
ERB2_HUMAN
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT ALA-1170.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS CYS-452; VAL-655 AND ALA-1170.
RA Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [5]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: Essential component of a neuregulin-receptor complex,
CC althought neuregulins do not interact with it alone. GP30 is a
CC potential ligand for this receptor. Not activated by EGF, TGF-

```
C       alpha and amphiregulin.
C    -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
C       tyrosine phosphate.
C    -!- SUBUNIT: Heterodimer with each of the other ERBB receptors
C       (Potential). Interacts with PRKCABP (By similarity).
C    -!- SUBCELLULAR LOCATION: Type I membrane protein.
C    -!- PTM: Ligand-binding increases phosphorylation on tyrosine
C       residues (By similarity).
C    -!- POLYMORPHISM: There are fours alleles due to the variations in
C       positions 654 and 655. Allele B1 (Ile-654/Ile-655) has a frequency
C       of 0.782; allele B2 (Ile-654/Val-655) has a frequency of 0.206;
C       allele B3 (Val-654/Val-655) has a frequency of 0.012.
C    -!- SIMILARITY: Belongs to the EGF receptor family.
C    ---------------------------------------------------------------------
C    This SWISS-PROT entry is copyright. It is produced through a collaboration
C    between the Swiss Institute of Bioinformatics and the EMBL outstation -
C    the European Bioinformatics Institute. There are no restrictions on its
C    use by non-profit institutions as long as its content is in no way
C    modified and this statement is not removed. Usage by and for commercial
C    entities requires a license agreement (See http://www.isb-sib.ch/announce/
C    or send an email to license@isb-sib.ch).
C    ---------------------------------------------------------------------
R    EMBL; M11767; AAA35808.1; -.
R    EMBL; M11761; AAA35808.1; JOINED.
R    EMBL; M11762; AAA35808.1; JOINED.
R    EMBL; M11763; AAA35808.1; JOINED.
R    EMBL; M11764; AAA35808.1; JOINED.
R    EMBL; M11765; AAA35808.1; JOINED.
R    EMBL; M11766; AAA35808.1; JOINED.
R    EMBL; M11730; AAA75493.1; -.
R    EMBL; M12036; AAA35978.1; -.
R    EMBL; AY208911; AAO18082.1; -.
R    EMBL; X03363; CAA27060.1; -.
R    PIR; A24571; A24571.
R    PDB; 1N8Z; 18-FEB-03.
R    PDB; 1QR1; 01-JAN-00.
R    Genew; HGNC:3430; ERBB2.
R    MIM; 164870; -.
R    GO; GO:0005012; F:Neu/ErbB-2 receptor activity; TAS.
R    GO; GO:0004716; F:receptor signaling protein tyrosine kinase . . .; TAS.
R    GO; GO:0008283; P:cell proliferation; TAS.
R    GO; GO:0007167; P:enzyme linked receptor protein signaling pa. . .; TAS.
R    GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.
R    GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
R    InterPro; IPR000494; EGFR_L_domain.
R    InterPro; IPR006211; Furin-like.
R    InterPro; IPR006212; Furin_repeat.
R    InterPro; IPR009030; Grow_fac_recep.
R    InterPro; IPR000719; Prot_kinase.
R    InterPro; IPR001245; Tyr_pkinase.
R    InterPro; IPR008266; Tyr_pkinase_AS.
R    InterPro; IPR004019; YLP_motif.
R    Pfam; PF00757; Furin-like; 1.
R    Pfam; PF00069; pkinase; 1.
R    Pfam; PF01030; Recep_L_domain; 2.
R    Pfam; PF02757; YLP; 2.
R    PRINTS; PR00109; TYRKINASE.
R    ProDom; PD000001; Prot_kinase; 1.
R    SMART; SM00261; FU; 4.
R    SMART; SM00219; TyrKc; 1.
R    PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
R    PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
R    PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
W    Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
W    Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
W    Polymorphism; 3D-structure.
T    SIGNAL        1     21       POTENTIAL.
T    CHAIN        22   1255       RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
T    DOMAIN       22    652       EXTRACELLULAR (POTENTIAL).
T    TRANSMEM    653    675       POTENTIAL.
T    DOMAIN      676   1255       CYTOPLASMIC (POTENTIAL).
T    DOMAIN      720    987       PROTEIN KINASE.
T    NP_BIND     726    734       ATP (BY SIMILARITY).
FT   BINDING     753    753       ATP (BY SIMILARITY).
FT   ACT_SITE    845    845       BY SIMILARITY.
FT   DISULFID    195    204       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    220    227       BY SIMILARITY.
FT   DISULFID    224    235       BY SIMILARITY.
FT   DISULFID    236    244       BY SIMILARITY.
FT   DISULFID    240    252       BY SIMILARITY.
FT   DISULFID    255    264       BY SIMILARITY.
FT   DISULFID    268    295       BY SIMILARITY.
FT   DISULFID    299    311       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    334    338       BY SIMILARITY.
FT   DISULFID    511    520       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    531    540       BY SIMILARITY.
FT   DISULFID    544    560       BY SIMILARITY.
FT   DISULFID    563    576       BY SIMILARITY.
FT   DISULFID    567    584       BY SIMILARITY.
FT   DISULFID    587    596       BY SIMILARITY.
FT   DISULFID    600    623       BY SIMILARITY.
FT   DISULFID    626    634       BY SIMILARITY.
FT   DISULFID    630    642       BY SIMILARITY.
FT   MOD_RES    1139   1139       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1248   1248       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    124    124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    530    530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    571    571       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     452    452       W -> C.
FT                                 /FTId=VAR_016317.
FT   VARIANT     654    654       I -> V (in allele B3; dbSNP:1801201).
FT                                 /FTId=VAR_004077.
FT   VARIANT     655    655       I -> V (in allele B2 and allele B3;
FT                                 dbSNP:1801200).
FT                                 /FTId=VAR_004078.
FT   VARIANT    1170   1170       P -> A.
FT                                 /FTId=VAR_016318.
SQ   SEQUENCE   1255 AA;  137909 MW;  39E9DFDA04DCF962 CRC64;

  Query Match             97.6%;  Score 41;  DB 1;  Length 1255;
  Best Local Similarity   88.9%;  Pred. No. 8.4;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              |||||||:
Db        665 VVLGVVFGI 673


RESULT 2
WSC3_YEAST
ID   WSC3_YEAST     STANDARD;      PRT;   556 AA.
AC   Q12215;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cell wall integrity and stress response component 3 precursor.
GN   WSC3 OR YOL105C OR HRE556.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96076631; PubMed=7502582;
RA   Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT   "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
RT   including the Tyl-H3 retrotransposon, the suf1(+) frameshift
RT   suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a
RT   delta element.";
```

```
L   Yeast 11:1069-1075(1995).
C   -!- SIMILARITY: Contains 1 WSC domain.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; Z48149; CAA88155.1; -.
R   EMBL; Z74847; CAA99123.1; -.
R   PIR; S51892; S51892.
R   GermOnline; 143527; -.
R   SGD; S0005465; WSC3.
R   GO; GO:0004888; F:transmembrane receptor activity; IGI.
R   GO; GO:0009408; P:response to heat; IGI.
R   GO; GO:0007266; P:Rho protein signal transduction; IGI.
R   InterPro; IPR002889; WSC.
R   Pfam; PF01822; WSC; 1.
R   SMART; SM00321; WSC; 1.
W   Cell wall; Transmembrane; Glycoprotein; Signal.
T   SIGNAL        1     38       POTENTIAL.
T   CHAIN         39    556      CELL WALL INTEGRITY AND STRESS RESPONSE
T                                COMPONENT 3.
T   DOMAIN        39    132      WSC.
T   DOMAIN       137    348      SER/THR-RICH.
T   TRANSMEM     385    405      POTENTIAL.
T   CARBOHYD      84     84      N-LINKED (GLCNAC. . .) (POTENTIAL).
T   CARBOHYD     367    367      N-LINKED (GLCNAC. . .) (POTENTIAL).
T   CARBOHYD     370    370      N-LINKED (GLCNAC. . .) (POTENTIAL).
T   CARBOHYD     473    473      N-LINKED (GLCNAC. . .) (POTENTIAL).
T   CARBOHYD     480    480      N-LINKED (GLCNAC. . .) (POTENTIAL).
Q   SEQUENCE   556 AA;  58229 MW;  DD37E277180001DA CRC64;

 Query Match             92.9%;  Score 39;  DB 1;  Length 556;
 Best Local Similarity   77.8%;  Pred. No. 10;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 VVLGVVFGV 9
             :|:||||||
b        385 IVIGVVFGV 393


ESULT 3
AV2_FUGRU
D   CAV2_FUGRU     STANDARD;      PRT;   162 AA.
C   Q9YGM9;
T   16-OCT-2001 (Rel. 40, Created)
T   16-OCT-2001 (Rel. 40, Last sequence update)
T   15-MAR-2004 (Rel. 43, Last annotation update)
E   Caveolin-2.
N   CAV2 OR CAV-2.
S   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
C   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
C   Tetradontoidea; Tetraodontidae; Takifugu.
X   NCBI_TaxID=31033;
N   [1]
P   SEQUENCE FROM N.A.
A   Cottage A.J.;
L   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
N   [2]
P   SEQUENCE FROM N.A.
X   MEDLINE=22799194; PubMed=12917688;
A   Thomas J.W., Touchman J.W., Blakesley R.W., Bouffard G.G.,
A   Beckstrom-Sternberg S.M., Margulies E.H., Blanchette M., Siepel A.C.,
A   Thomas P.J., McDowell J.C., Maskeri B., Hansen N.F., Schwartz M.S.,
A   Weber R.J., Kent W.J., Karolchik D., Bruen T.C., Bevan R.,
A   Cutler D.J., Schwartz S., Elnitski L., Idol J.R., Prasad A.B.,
A   Lee-Lin S.-Q., Maduro V.V., Summers T.J., Portnoy M.E., Dietrich N.L.,
A   Akhter N., Ayele K., Benjamin B., Cariaga K., Brinkley C.P.,
A   Brooks S.Y., Granite S., Guan X., Gupta J., Haghighi P., Ho S.-L.,
A   Huang M.C., Karlins E., Laric P.L., Legaspi R., Lim M.J., Maduro Q.L.,
A   Masiello C.A., Mastrian S.D., McCloskey J.C., Pearson R.,
A   Stantripop S., Tiongson E.E., Tran J.T., Tsurgeon C., Vogt J.L.,
A   Walker M.A., Wetherby K.D., Wiggins L.S., Young A.C., Zhang L.-H.,
A   Osoegawa K., Zhu B., Zhao B., Shu C.L., De Jong P.J., Lawrence C.E.,
A   Smit A.F., Chakravarti A., Haussler D., Green P., Miller W.,
A   Green E.D.;
T   "Comparative analyses of multi-species sequences from targeted
T   genomic regions.";
L   Nature 424:788-793(2003).
C   -!- FUNCTION: May act as a scaffolding protein within caveolar
C       membranes. Interacts directly with G-protein alpha subunits and
C       can functionally regulate their activity (By similarity).
C   -!- SUBUNIT: Homooligomer (By similarity).
C   -!- SUBCELLULAR LOCATION: Membrane protein of caveolae. Potential
C       hairpin-like structure in the membrane (By similarity).
C   -!- SIMILARITY: Belongs to the caveolin family.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; AJ010316; CAA09081.1; -.
R   EMBL; AC090119; AAL40363.1; -.
R   InterPro; IPR001612; Caveolin.
R   Pfam; PF01146; Caveolin; 1.
R   PROSITE; PS01210; CAVEOLIN; 1.
K   Transmembrane; Lipoprotein.
T   DOMAIN        1     86       CYTOPLASMIC (POTENTIAL).
T   TRANSMEM     87    107       POTENTIAL.
T   DOMAIN      108    162       CYTOPLASMIC (POTENTIAL).
Q   SEQUENCE   162 AA;  18236 MW;  1D7CF4907D491253 CRC64;

  Query Match             88.1%;  Score 37;  DB 1;  Length 162;
  Best Local Similarity   87.5%;  Pred. No. 8.7;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         2 VLGVVFGV 9
             :|||||||
Db        99 ILGVVFGV 106


RESULT 4
CYS5_TOBAC
ID   CYS5_TOBAC     STANDARD;      PRT;   135 AA.
AC   P49099;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Cytochrome b5, seed isoform.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Samsun; TISSUE=Seed;
RX   MEDLINE=96009039; PubMed=7580860;
RA   Napier J.A., Smith M.A., Stobart A.K., Shewry P.R.;
RT   "Isolation of a cDNA encoding a cytochrome b5 specifically expressed
RT   in developing tobacco seeds.";
RL   Planta 197:200-202(1995).
CC   -!- FUNCTION: Cytochrome b5 is a membrane bound hemoprotein which
CC       function as an electron carrier for several membrane bound
CC       oxygenases. May play a key role in the modification by
CC       desaturation of fatty acids in the endoplasmic reticulum, which in
```

```
CC       the developing seed is utilized for membrane synthesis and in the
CC       developmentally regulated production of large amounts of storage
CC       lipids.
CC   -!- SUBCELLULAR LOCATION: Microsomal membrane. Bound to the
CC       cytoplasmic side of the endoplasmic reticulum (By similarity).
CC   -!- TISSUE SPECIFICITY: Specifically expressed in developing seeds.
CC   -!- SIMILARITY: Belongs to the cytochrome b5 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80008; CAA56318.1; -.
DR   PIR; S49200; S49200.
DR   HSSP; P00171; 1CYO.
DR   InterPro; IPR001199; Cyt_B5.
DR   Pfam; PF00173; heme_1; 1.
DR   PRINTS; PR00363; CYTOCHROMEB5.
DR   ProDom; PD000612; Cyt_B5; 1.
DR   PROSITE; PS00191; CYTOCHROME_B5_1; 1.
DR   PROSITE; PS50255; CYTOCHROME_B5_2; 1.
KW   Electron transport; Transmembrane; Heme; Iron; Microsome;
KW   Multigene family.
FT   TRANSMEM    107    127       POTENTIAL.
FT   METAL        40     40       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   METAL        64     64       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   135 AA;  14869 MW;  A36CCA081A72ECBC CRC64;

  Query Match             85.7%;  Score 36;  DB 1;  Length 135;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::||| |||
Db        119 IILGVAFGV 127


RESULT 5
CT24_HUMAN
ID   CT24_HUMAN     STANDARD;      PRT;   137 AA.
AC   Q9BUV8; O00605; Q9BT03; Q9BZU7; Q9UI05;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein C20orf24 (Rab5-interacting protein) (RIP5) (PNAS-11).
GN   C20ORF24.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Adrenal gland;
RA   Jiang C., Zhang C., Huang C., Peng Y., Gu Y., Zhang L., Wu T., Li Y.,
RA   Han Z., Wang Y., Chen Z., Fu G.;
RT   "A novel gene expressed in human adrenal gland.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RA   Yu W.-Q., Sun B.-Z., Chai Y.-B., Zhu F., Liu X.-S., Li Z., Lu F.,
RA   Yan W., Yang H., Zhao Z.-L.;
RT   "Human acute promyelocytic leukemia cell line NB4's apoptosis related
RT   genes.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).
RC   TISSUE=Lung, and Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 15-137 FROM N.A. (ISOFORM 1).
RX   MEDLINE=96421776; PubMed=8824393;
RA   Vitale G., Alexandrov K., Ullrich O., Horiuchi H., Giner A.,
RA   Dobson C., Baykova O., Gournier H., Stenmark H., Zerial M.;
RT   "The GDP/GTP cycle of Rab5 in the regulation of endocytotic membrane
RT   traffic.";
RL   Cold Spring Harb. Symp. Quant. Biol. 60:211-220(1995).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=1;
CC         IsoId=Q9BUV8-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9BUV8-2; Sequence=VSP_003796;
CC       Name=3;
CC         IsoId=Q9BUV8-3; Sequence=VSP_003797;
CC       Name=4;
CC         IsoId=Q9BUV8-4; Sequence=VSP_003795;
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------
R   EMBL; AF112213; AAF17201.1; -.
R   EMBL; AF274936; AAK07515.1; -.
R   EMBL; AL050318; CAB75367.1; -.
R   EMBL; BC001871; AAH01871.1; -.
R   EMBL; BC004446; AAH04446.1; -.
R   EMBL; S83364; AAB50849.1; -.
R   Genew; HGNC:15870; C20orf24.
W   Alternative splicing.
T   VARSPLIC     74    137       FCLINAGVLYLYFSNYLQIDEEEYGGTWELTKEGFMTSFAL
T                                FMVCVADSFTTGHLDHLLLHCHPL -> KGHCCSGAVCVCDD
T                                (in isoform 4).
T                                /FTId=VSP_003795.
T   VARSPLIC    118    137       CVADSFTTGHLDHLLLHCHPL -> IWIIFYTAIHYD (in
T                                isoform 2).
T                                /FTId=VSP_003796.
T   VARSPLIC    118    127       CVADSFTTGH -> I (in isoform 3).
T                                /FTId=VSP_003797.
T   CONFLICT     15     33       LANGALKVSVWSKVLRSDA -> PWANGASKSPSGVRCCGR
T                                R (IN REF. 5).
T   CONFLICT     95    100       EEYGGT -> GRILVAR (IN REF. 2).
T   CONFLICT    100    100       T -> Q (IN REF. 5).
T   CONFLICT    110    110       T -> P (IN REF. 5).
T   CONFLICT    114    137       LFMVCVADSFTTGHLDHLLLHCHPL -> IVHGHLLLFTSHP
T                                YSMMVSDSK (IN REF. 5).
F   CONFLICT    137    137       L -> FIT (IN REF. 2).
Q   SEQUENCE   137 AA;  15487 MW;  4DC7A8E5D61B4C57 CRC64;

 Query Match             85.7%;  Score 36;  DB 1;  Length 137;
 Best Local Similarity   77.8%;  Pred. No. 12;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Y          1 VVLGVVFGV 9
             |||||::||
D         54 VVLGVIWGV 62


ESULT 6
YB5_TOBAC
D   CYB5_TOBAC     STANDARD;      PRT;   136 AA.
C   P49098;
T   01-FEB-1996 (Rel. 33, Created)
T   01-FEB-1996 (Rel. 33, Last sequence update)
T   15-MAR-2004 (Rel. 43, Last annotation update)
E   Cytochrome b5.
S   Nicotiana tabacum (Common tobacco).
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C   lamiids; Solanales; Solanaceae; Nicotiana.
K   NCBI_TaxID=4097;
N   [1]
P   SEQUENCE FROM N.A.
C   TISSUE=Leaf;
K   MEDLINE=94325476; PubMed=8049375;
A   Smith M.A., Stobart A.K., Shewry P.R., Napier J.A.;
T   "Tobacco cytochrome b5: cDNA isolation, expression analysis and in
T   vitro protein targeting.";
L   Plant Mol. Biol. 25:527-537(1994).
C   -!- FUNCTION: Cytochrome b5 is a membrane bound hemoprotein which
C       function as an electron carrier for several membrane bound
C       oxygenases. May play a key role in the modification by
C       desaturation of fatty acids in the endoplasmic reticulum, which in
C       the developing seed is utilized for membrane synthesis and in the
C       developmentally regulated production of large amounts of storage
C       lipids. Is involved in the reduction of cytochrome P-450 and
C       may therefore be involved in flavonoid biosynthesis in the petals.
C   -!- SUBCELLULAR LOCATION: Microsomal membrane. Bound to the
C       cytoplasmic side of the endoplasmic reticulum (By similarity).
C   -!- TISSUE SPECIFICITY: Is highly expressed in developing seeds,
CC      moderately expressed in flowers, and is expressed at low
CC      levels in the leaf.
CC  -!- SIMILARITY: Belongs to the cytochrome b5 family.
CC  ----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ----------------------------------------------------------------------
DR  EMBL; X71441; CAA50575.1; ALT_INIT.
DR  EMBL; X68140; CAA48240.1; -.
DR  HSSP; P04166; 1B5M.
DR  InterPro; IPR001199; Cyt_B5.
DR  Pfam; PF00173; heme_1; 1.
DR  PRINTS; PR00363; CYTOCHROMEB5.
DR  ProDom; PD000612; Cyt_B5; 1.
DR  PROSITE; PS00191; CYTOCHROME_B5_1; 1.
DR  PROSITE; PS50255; CYTOCHROME_B5_2; 1.
KW  Electron transport; Transmembrane; Heme; Iron; Microsome;
KW  Multigene family.
FT  TRANSMEM    107    127       POTENTIAL.
FT  METAL        40     40       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT  METAL        64     64       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT  CONFLICT     10     11       LA -> EF (IN REF. 1; CAA48240).
FT  CONFLICT    105    105       MISSING (IN REF. 1; CAA48240).
SQ  SEQUENCE   136 AA;  14979 MW;  DACE9EA695B2835F CRC64;

  Query Match             83.3%;  Score 35;  DB 1;  Length 136;
  Best Local Similarity   55.6%;  Pred. No. 17;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::|||  ||:
Db        119 IILGVAFGI 127


RESULT 7
COBS_COREF
ID  COBS_COREF     STANDARD;      PRT;   275 AA.
AC  Q8FNQ1;
DT  10-OCT-2003 (Rel. 42, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Cobalamin synthase (EC 2.-.-.-).
GN  COBS OR CE2093.
OS  Corynebacterium efficiens.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX  NCBI_TaxID=152794;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX  MEDLINE=22723752; PubMed=12840036;
RA  Nishio Y., Nakamura Y., Kawarabayasi Y., Usuda Y., Kimura E.,
RA  Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA  Gojobori T.;
RT  "Comparative complete genome sequence analysis of the amino acid
RT  replacements responsible for the thermostability of Corynebacterium
RT  efficiens.";
RL  Genome Res. 13:1572-1579(2003).
CC  -!- FUNCTION: Joins Ado-cobinamide-GDP and alpha-ribazole to generate
CC      adenosylcobalamin (Ado-cobalamin) (By similarity).
CC  -!- CATALYTIC ACTIVITY: GDP-cobinamide + alpha-ribazole = cobalamin +
CC      GMP.
CC  -!- PATHWAY: Cobalamin biosynthesis; last step.
CC  -!- SIMILARITY: Belongs to the cobS family.
CC  ----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
```

```
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------
R   EMBL; AP005221; BAC18903.1; -.
R   HAMAP; MF_00719; -; 1.
R   InterPro; IPR003805; CobS_synth.
R   Pfam; PF02654; CobS; 1.
W   Cobalamin biosynthesis; Transferase; Complete proteome.
Q   SEQUENCE   275 AA;  27958 MW;  36E8A1625EE7B3EB CRC64;

 Query Match             83.3%;  Score 35;  DB 1;  Length 275;
 Best Local Similarity   87.5%;  Pred. No. 30;
 Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

y          2 VLGVVFGV 9
             |:||||||
b         52 VVGVVFGV 59


ESULT 8
LX_SHIFL
D   ALX_SHIFL      STANDARD;      PRT;   320 AA.
C   Q83Q35; Q7UBI0;
T   15-MAR-2004 (Rel. 43, Created)
T   15-MAR-2004 (Rel. 43, Last sequence update)
T   15-MAR-2004 (Rel. 43, Last annotation update)
E   Alx protein.
N   ALX OR SF3128 OR S3335.
S   Shigella flexneri.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C   Enterobacteriaceae; Shigella.
X   NCBI_TaxID=623;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=301 / Serotype 2a;
X   MEDLINE=22272406; PubMed=12384590;
A   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
A   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
A   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
A   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
A   Yu J.;
T   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
T   through comparison with genomes of Escherichia coli K12 and O157.";
L   Nucleic Acids Res. 30:4432-4441(2002).
N   [2]
P   SEQUENCE FROM N.A.
C   STRAIN=2457T / ATCC 700930 / Serotype 2a;
X   MEDLINE=22590274; PubMed=12704152;
A   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
A   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
A   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
A   Schwartz D.C., Blattner F.R.;
T   "Complete genome sequence and comparative genomics of Shigella
T   flexneri serotype 2a strain 2457T.";
L   Infect. Immun. 71:2775-2786(2003).
C   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
C   -!- SIMILARITY: Belongs to the terC family.
C   ----------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ----------------------------------------------------------------------
R   EMBL; AE015324; AAN44600.1; -.
R   EMBL; AE016989; AAP18413.1; -.
R   InterPro; IPR005496; TerC.
R   Pfam; PF03741; TerC; 1.
W   Transmembrane; Complete proteome.
FT   TRANSMEM      7     27       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    114    134       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    199    219       POTENTIAL.
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    262    282       POTENTIAL.
FT   TRANSMEM    287    307       POTENTIAL.
FT   CONFLICT    320    320       G -> GE (IN REF. 2).
SQ   SEQUENCE   320 AA;  35852 MW;  31C173442799C384 CRC64;

  Query Match             83.3%;  Score 35;  DB 1;  Length 320;
  Best Local Similarity   77.8%;  Pred. No. 34;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              | ||||||:
Db        292 VSLGVVFGI 300


RESULT 9
ALX_ECO57
ID   ALX_ECO57      STANDARD;      PRT;   321 AA.
AC   Q8XAJ0;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Alx protein.
GN   ALX OR Z4441 OR ECS3970.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the terC family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005538; AAG58221.1; -.
DR   EMBL; AP002564; BAB37393.1; -.
DR   PIR; A85970; A85970.
DR   PIR; B91125; B91125.
DR   InterPro; IPR005496; TerC.
```

```
R   Pfam; PF03741; TerC; 1.
W   Transmembrane; Complete proteome.
T   TRANSMEM      7     27       POTENTIAL.
T   TRANSMEM     44     64       POTENTIAL.
T   TRANSMEM     90    110       POTENTIAL.
T   TRANSMEM    114    134       POTENTIAL.
T   TRANSMEM    136    156       POTENTIAL.
T   TRANSMEM    199    219       POTENTIAL.
T   TRANSMEM    226    246       POTENTIAL.
T   TRANSMEM    262    282       POTENTIAL.
T   TRANSMEM    287    307       POTENTIAL.
Q   SEQUENCE   321 AA;  35951 MW;  B6FB7173442799C3 CRC64;

 Query Match             83.3%;  Score 35;  DB 1;  Length 321;
 Best Local Similarity   77.8%;  Pred. No. 34;
 Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

y          1 VVLGVVFGV 9
             | ||||||:
b        292 VSLGVVFGI 300


ESULT 10
LX_ECOL6
D   ALX_ECOL6      STANDARD;      PRT;   321 AA.
C   Q8FDE1;
T   15-MAR-2004 (Rel. 43, Created)
T   15-MAR-2004 (Rel. 43, Last sequence update)
T   15-MAR-2004 (Rel. 43, Last annotation update)
E   Alx protein.
N   ALX OR C3846.
S   Escherichia coli O6.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C   Enterobacteriaceae; Escherichia.
X   NCBI_TaxID=217992;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=O6:H1 / CFT073 / ATCC 700928;
X   MEDLINE=22388234; PubMed=12471157;
A   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
A   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
A   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
A   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
T   "Extensive mosaic structure revealed by the complete genome sequence
T   of uropathogenic Escherichia coli.";
L   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
C   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
C   -!- SIMILARITY: Belongs to the terC family.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; AE016767; AAN82291.1; -.
R   InterPro; IPR005496; TerC.
R   Pfam; PF03741; TerC; 1.
W   Transmembrane; Complete proteome.
T   TRANSMEM      7     27       POTENTIAL.
T   TRANSMEM     44     64       POTENTIAL.
T   TRANSMEM     90    110       POTENTIAL.
T   TRANSMEM    114    134       POTENTIAL.
T   TRANSMEM    136    156       POTENTIAL.
T   TRANSMEM    199    219       POTENTIAL.
T   TRANSMEM    226    246       POTENTIAL.
T   TRANSMEM    262    282       POTENTIAL.
T   TRANSMEM    287    307       POTENTIAL.
)   SEQUENCE   321 AA;  36053 MW;  A9EB77A091279662 CRC64;

 Query Match             83.3%;  Score 35;  DB 1;  Length 321;
   Best Local Similarity   77.8%;  Pred. No. 34;
   Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

 Qy          1 VVLGVVFGV 9
               | ||||||:
 Db        292 VSLGVVFGI 300


 RESULT 11
 ALX_ECOLI
 ID   ALX_ECOLI      STANDARD;      PRT;   321 AA.
 AC   P42601;
 DT   01-NOV-1995 (Rel. 32, Created)
 DT   01-NOV-1995 (Rel. 32, Last sequence update)
 DT   15-MAR-2004 (Rel. 43, Last annotation update)
 DE   Alx protein.
 GN   ALX OR B3088.
 OS   Escherichia coli.
 OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC   Enterobacteriaceae; Escherichia.
 OX   NCBI_TaxID=562;
 RN   [1]
 RP   SEQUENCE FROM N.A.
 RC   STRAIN=K12 / MG1655;
 RX   MEDLINE=97426617; PubMed=9278503;
 RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA   Mau B., Shao Y.;
 RT   "The complete genome sequence of Escherichia coli K-12.";
 RL   Science 277:1453-1474(1997).
 RN   [2]
 RP   GENE NAME, AND INDUCTION.
 RC   STRAIN=K12;
 RX   MEDLINE=90202745; PubMed=2108134;
 RA   Bingham R.J., Hall K.S., Slonczewski J.L.;
 RT   "Alkaline induction of a novel gene locus, alx, in Escherichia coli.";
 RL   J. Bacteriol. 172:2184-2186(1990).
 RN   [3]
 RP   INDUCTION.
 RC   STRAIN=K12;
 RX   MEDLINE=22103114; PubMed=12107143;
 RA   Stancik L.M., Stancik D.M., Schmidt B., Barnhart D.M., Yoncheva Y.N.,
 RA   Slonczewski J.L.;
 RT   "pH-dependent expression of periplasmic proteins and amino acid
 RT   catabolism in Escherichia coli.";
 RL   J. Bacteriol. 184:4246-4258(2002).
 CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC   -!- INDUCTION: By extreme alkaline conditions.
 CC   -!- SIMILARITY: Belongs to the terC family.
 CC   ---------------------------------------------------------------------
 CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC   the European Bioinformatics Institute. There are no restrictions on its
 CC   use by non-profit institutions as long as its content is in no way
 CC   modified and this statement is not removed. Usage by and for commercial
 CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC   or send an email to license@isb-sib.ch).
 CC   ---------------------------------------------------------------------
 DR   EMBL; U18997; AAA57890.1; -.
 DR   EMBL; AE000391; AAC76123.1; -.
 DR   PIR; E65097; E65097.
 DR   EcoGene; EG12731; alx.
 DR   InterPro; IPR005496; TerC.
 DR   Pfam; PF03741; TerC; 1.
 KW   Transmembrane; Complete proteome.
 FT   TRANSMEM      7     27       POTENTIAL.
 FT   TRANSMEM     44     64       POTENTIAL.
 FT   TRANSMEM     90    110       POTENTIAL.
 FT   TRANSMEM    114    134       POTENTIAL.
 FT   TRANSMEM    136    156       POTENTIAL.
 FT   TRANSMEM    199    219       POTENTIAL.
 FT   TRANSMEM    226    246       POTENTIAL.
```

```
T   TRANSMEM    262    282       POTENTIAL.
T   TRANSMEM    287    307       POTENTIAL.
Q   SEQUENCE   321 AA;  35837 MW;  726A56FD23F38C4E CRC64;

 Query Match             83.3%;  Score 35;  DB 1;  Length 321;
 Best Local Similarity   77.8%;  Pred. No. 34;
 Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

y          1 VVLGVVFGV 9
              | ||||||:
b        292 VSLGVVFGI 300


ESULT 12
GCC_ECOLI
D   SGCC_ECOLI     STANDARD;      PRT;   437 AA.
C   P39365;
T   01-FEB-1995 (Rel. 31, Created)
T   01-FEB-1995 (Rel. 31, Last sequence update)
T   28-FEB-2003 (Rel. 41, Last annotation update)
E   Putative phosphotransferase enzyme II, C component sgcC.
N   SGCC OR B4304.
S   Escherichia coli.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C   Enterobacteriaceae; Escherichia.
X   NCBI_TaxID=562;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=K12 / MG1655;
X   MEDLINE=95334362; PubMed=7610040;
A   Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
A   Blattner F.R.;
T   "Analysis of the Escherichia coli genome VI: DNA sequence of the
T   region from 92.8 through 100 minutes.";
L   Nucleic Acids Res. 23:2105-2119(1995).
N   [2]
P   DISCUSSION OF SEQUENCE.
A   Reizer J., Charbit A., Reizer A., Saier M.H. Jr.;
T   "Novel phosphotransferases system genes revealed by bacterial genome
T   analysis: operons encoding homologues of sugar-specific permease
T   domains of the phosphotransferase system and pentose catabolic
T   enzymes.";
L   Genome Sci. Technol. 1:53-75(1996).
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
C       (Potential).
C   -!- SIMILARITY: Contains 1 PTS EIIC domain.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; U14003; AAA97200.1; -.
R   EMBL; AE000501; AAC77260.1; -.
R   PIR; S56529; S56529.
R   EcoGene; EG12556; sgcC.
R   InterPro; IPR004703; Gal_spec_IIC.
R   Pfam; PF03611; EIIC-GAT; 1.
R   TIGRFAMs; TIGR00827; EIIC-GAT; 1.
W   Phosphotransferase system; Sugar transport; Transmembrane;
W   Inner membrane; Complete proteome.
T   TRANSMEM      5     25       POTENTIAL.
T   TRANSMEM     35     55       POTENTIAL.
T   TRANSMEM     88    108       POTENTIAL.
T   TRANSMEM    134    154       POTENTIAL.
T   TRANSMEM    173    193       POTENTIAL.
T   TRANSMEM    215    235       POTENTIAL.
T   TRANSMEM    236    256       POTENTIAL.
T   TRANSMEM    302    322       POTENTIAL.
T   TRANSMEM    325    345       POTENTIAL.
FT  TRANSMEM    354    374       POTENTIAL.
FT  TRANSMEM    385    405       POTENTIAL.
FT  TRANSMEM    410    430       POTENTIAL.
SQ  SEQUENCE   437 AA;  46684 MW;  5FE8C034BD925F21 CRC64;

  Query Match             83.3%;  Score 35;  DB 1;  Length 437;
  Best Local Similarity   66.7%;  Pred. No. 43;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ||||::||:
Db        225 VVLGLIFGL 233


RESULT 13
CLCA_ECO57
ID  CLCA_ECO57     STANDARD;      PRT;   473 AA.
AC  P58244;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Voltage-gated ClC-type chloride channel clcA.
GN  CLCA OR ERIC OR Z0166 OR ECS0159.
OS  Escherichia coli O157:H7.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Escherichia.
OX  NCBI_TaxID=83334;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX  MEDLINE=21074935; PubMed=11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA  Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533(2001).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
CC  -!- FUNCTION: Probably acts as an electrical shunt for an outwardly-
CC      directed proton pump that is linked to amino acid decarboxylation,
CC      as part of the extreme acid resistance (XAR) response (By
CC      similarity).
CC  -!- SUBUNIT: Homodimer (By similarity).
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC      (Probable).
CC  -!- DOMAIN: Helix R might transduce intracellular events into channel
CC      gating (By similarity).
CC  -!- MISCELLANEOUS: The dimeric channel has a two-fold axis
CC      perpendicular to the membrane plane; each of the subunits within
CC      the dimer exhibits an antiparallel architecture and forms its own
CC      ion-conducting pore. The channel is probably activated by chloride
CC      ions, which appear to exert this gating effect by actually
CC      entering the pore. The ion conduction and gating are thus closely
CC      linked (By similarity).
CC  -!- MISCELLANEOUS: The two ClC channels in this bacterium, clcA and
CC      clcB, act redundantly (By similarity).
CC  -!- SIMILARITY: Belongs to the chloride channel family.
CC  ---------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE005192; AAG54459.1; -.
DR   EMBL; AP002550; BAB33582.1; -.
DR   PIR; G90648; G90648.
DR   HSSP; P37019; 1KPK.
DR   HAMAP; MF_01128; -; 1.
DR   InterPro; IPR001807; Cl-channel_volt.
DR   Pfam; PF00654; voltage_CLC; 1.
DR   PRINTS; PR00762; CLCHANNEL.
KW   Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW   Chloride channel; Chloride; Inner membrane; Transmembrane;
KW   Complete proteome.
FT   DOMAIN        1     32       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     33     65       BY SIMILARITY.
FT   DOMAIN       66     78       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM     79    100       BY SIMILARITY.
FT   DOMAIN      106    110       SELECTIVITY FILTER PART_1 (BY
FT                                SIMILARITY).
FT   DOMAIN      109    116       IN-MEMBRANE HELIX (BY SIMILARITY).
FT   DOMAIN      117    126       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    127    140       BY SIMILARITY.
FT   DOMAIN      146    150       SELECTIVITY FILTER PART_2 (BY
FT                                SIMILARITY).
FT   TRANSMEM    148    164       BY SIMILARITY.
FT   DOMAIN      165    170       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    171    189       BY SIMILARITY.
FT   TRANSMEM    190    192       LOOP BETWEEN TWO HELICES (BY SIMILARITY).
FT   TRANSMEM    193    202       BY SIMILARITY.
FT   DOMAIN      203    214       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    215    232       BY SIMILARITY.
FT   DOMAIN      233    252       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    253    284       BY SIMILARITY.
FT   DOMAIN      285    287       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    288    307       BY SIMILARITY.
FT   DOMAIN      308    329       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    330    349       BY SIMILARITY.
FT   DOMAIN      355    359       SELECTIVITY FILTER PART_3 (BY
FT                                SIMILARITY).
FT   TRANSMEM    357    378       BY SIMILARITY.
FT   DOMAIN      379    386       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    387    401       BY SIMILARITY.
FT   TRANSMEM    402    404       LOOP BETWEEN TWO HELICES (BY SIMILARITY).
FT   TRANSMEM    405    416       BY SIMILARITY.
FT   TRANSMEM    417    421       LOOP BETWEEN TWO HELICES (BY SIMILARITY).
FT   TRANSMEM    422    438       BY SIMILARITY.
FT   DOMAIN      439    473       CYTOPLASMIC (BY SIMILARITY).
FT   SITE        148    148       MAY BE INVOLVED IN GATING (BY
FT                                SIMILARITY).
FT   BINDING     107    107       CHLORIDE (BY SIMILARITY).
FT   BINDING     356    356       CHLORIDE (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     357    357       CHLORIDE (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     445    445       CHLORIDE (BY SIMILARITY).
SQ   SEQUENCE   473 AA;  50334 MW;  72255396D9676B23 CRC64;

 Query Match             83.3%;  Score 35;  DB 1;  Length 473;
 Best Local Similarity   44.4%;  Pred. No. 46;
 Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::||::||:
Db        256 LILGIIFGI 264


RESULT 14
CLCA_ECOL6
ID   CLCA_ECOL6     STANDARD;      PRT;   473 AA.
AC   Q8FL15;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Voltage-gated ClC-type chloride channel clcA.
GN   CLCA OR ERIC OR C0190.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC   -!- FUNCTION: Probably acts as an electrical shunt for an outwardly-
CC       directed proton pump that is linked to amino acid decarboxylation,
CC       as part of the extreme acid resistance (XAR) response (By
CC       similarity).
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- DOMAIN: Helix R might transduce intracellular events into channel
CC       gating (By similarity).
CC   -!- MISCELLANEOUS: The dimeric channel has a two-fold axis
CC       perpendicular to the membrane plane; each of the subunits within
CC       the dimer exhibits an antiparallel architecture and forms its own
CC       ion-conducting pore. The channel is probably activated by chloride
CC       ions, which appear to exert this gating effect by actually
CC       entering the pore. The ion conduction and gating are thus closely
CC       linked (By similarity).
CC   -!- MISCELLANEOUS: The two ClC channels in this bacterium, clcA and
CC       clcB, act redundantly (By similarity).
CC   -!- SIMILARITY: Belongs to the chloride channel family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE016755; AAN78684.1; -.
DR   HAMAP; MF_01128; -; 1.
DR   InterPro; IPR001807; Cl-channel_volt.
DR   Pfam; PF00654; voltage_CLC; 1.
DR   PRINTS; PR00762; CLCHANNEL.
KW   Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW   Chloride channel; Chloride; Inner membrane; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     31     53       POTENTIAL.
FT   TRANSMEM     75     97       POTENTIAL.
FT   TRANSMEM    124    146       POTENTIAL.
FT   TRANSMEM    179    201       POTENTIAL.
FT   TRANSMEM    210    232       POTENTIAL.
FT   TRANSMEM    252    274       POTENTIAL.
FT   TRANSMEM    287    309       POTENTIAL.
FT   TRANSMEM    319    341       POTENTIAL.
FT   TRANSMEM    354    376       POTENTIAL.
FT   TRANSMEM    391    413       POTENTIAL.
FT   TRANSMEM    418    440       POTENTIAL.
FT   SITE        148    148       MAY BE INVOLVED IN GATING (BY
FT                                SIMILARITY).
FT   BINDING     107    107       CHLORIDE (BY SIMILARITY).
FT   BINDING     356    356       CHLORIDE (VIA AMIDE NITROGEN) (BY
FT                                SIMILARITY).
FT   BINDING     357    357       CHLORIDE (VIA AMIDE NITROGEN) (BY
```

```
FT                              SIMILARITY).
FT   BINDING     445    445     CHLORIDE (BY SIMILARITY).
SQ   SEQUENCE   473 AA;  50403 MW;  7225539769676B23 CRC64;

  Query Match             83.3%;  Score 35;  DB 1;  Length 473;
  Best Local Similarity   44.4%;  Pred. No. 46;
  Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::||::||:
Db        256 LILGIIFGI 264


RESULT 15
CLCA_ECOLI
ID   CLCA_ECOLI     STANDARD;      PRT;   473 AA.
AC   P37019; P77394;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Voltage-gated ClC-type chloride channel clcA.
GN   CLCA OR ERIC OR B0155.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RX   MEDLINE=94261430; PubMed=8202364;
RA   Fujita N., Mori H., Yura T., Ishihama A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 2.4-4.1 min (110,917-193,643 bp) region.";
RL   Nucleic Acids Res. 22:1637-1639(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA   Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA   Davis R.W.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   CHARACTERIZATION.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=20009653; PubMed=10539975;
RA   Maduke M., Pheasant D.J., Miller C.;
RT   "High-level expression, functional reconstitution, and quaternary
RT   structure of a prokaryotic ClC-type chloride channel.";
RL   J. Gen. Physiol. 114:713-722(1999).
RN   [5]
RP   CHARACTERIZATION.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=20115456; PubMed=10648805;
RA   Purdy M.D., Wiener M.C.;
RT   "Expression, purification, and initial structural characterization of
RT   YadQ, a bacterial homolog of mammalian ClC chloride channel
RT   proteins.";
RL   FEBS Lett. 466:26-28(2000).
RN   [6]
RP   FUNCTION.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=22272680; PubMed=12384697;
RA   Iyer R., Iverson T.M., Accardi A., Miller C.;
RT   "A biological role for prokaryotic ClC chloride channels.";
RL   Nature 419:715-718(2002).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (6.5 ANGSTROMS).
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=21037970; PubMed=11196649;
RA   Mindell J.A., Maduke M., Miller C., Grigorieff N.;
RT   "Projection structure of a ClC-type chloride channel at 6.5 A
RT   resolution.";
RL   Nature 409:219-223(2001).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
RX   MEDLINE=21655566; PubMed=11796999;
RA   Dutzler R., Campbell E.B., Cadene M., Chait B.T., MacKinnon R.;
RT   "X-ray structure of a ClC chloride channel at 3.0 A reveals the
RT   molecular basis of anion selectivity.";
RL   Nature 415:287-294(2002).
CC   -!- FUNCTION: Probably acts as an electrical shunt for an outwardly-
CC       directed proton pump that is linked to amino acid decarboxylation,
CC       as part of the extreme acid resistance (XAR) response.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- INDUCTION: By acid-shock conditions.
CC   -!- DOMAIN: Helix R might transduce intracellular events into channel
CC       gating.
CC   -!- MISCELLANEOUS: The dimeric channel has a two-fold axis
CC       perpendicular to the membrane plane; each of the subunits within
CC       the dimer exhibits an antiparallel architecture and forms its own
CC       ion-conducting pore. The channel is probably activated by chloride
CC       ions, which appear to exert this gating effect by actually
CC       entering the pore. The ion conduction and gating are thus closely
CC       linked.
CC   -!- MISCELLANEOUS: The two ClC channels in this bacterium, clcA and
CC       clcB, act redundantly.
CC   -!- SIMILARITY: Belongs to the chloride channel family.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 11.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26562; BAB96732.1; ALT_FRAME.
DR   EMBL; AE000125; AAC73266.1; -.
DR   EMBL; U70214; AAB08585.1; -.
DR   PIR; C64739; C64739.
DR   PDB; 1KPK; 23-JAN-02.
DR   EcoGene; EG12331; clcA.
DR   HAMAP; MF_01128; -; 1.
DR   InterPro; IPR001807; Cl-channel_volt.
DR   Pfam; PF00654; voltage_CLC; 1.
DR   PRINTS; PR00762; CLCHANNEL.
KW   Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW   Chloride channel; Chloride; Inner membrane; Transmembrane;
KW   3D-structure; Complete proteome.
FT   DOMAIN        1     32       CYTOPLASMIC.
FT   TRANSMEM     33     65
FT   DOMAIN       66     78       EXTRACELLULAR.
FT   TRANSMEM     79    100
FT   DOMAIN      106    110       SELECTIVITY FILTER PART_1.
FT   DOMAIN      109    116       IN-MEMBRANE HELIX.
FT   DOMAIN      117    126       CYTOPLASMIC.
FT   TRANSMEM    127    140
FT   DOMAIN      146    150       SELECTIVITY FILTER PART_2.
FT   TRANSMEM    148    164
FT   DOMAIN      165    170       CYTOPLASMIC.
FT   TRANSMEM    171    189
FT   TRANSMEM    190    192       LOOP BETWEEN TWO HELICES.
```

T	TRANSMEM	193	202	
T	DOMAIN	203	214	CYTOPLASMIC.
T	TRANSMEM	215	232	
T	DOMAIN	233	252	EXTRACELLULAR.
T	TRANSMEM	253	284	
T	DOMAIN	285	287	CYTOPLASMIC.
T	TRANSMEM	288	307	
T	DOMAIN	308	329	EXTRACELLULAR.
T	TRANSMEM	330	349	
T	DOMAIN	355	359	SELECTIVITY FILTER PART_3.
T	TRANSMEM	357	378	
T	DOMAIN	379	386	EXTRACELLULAR.
T	TRANSMEM	387	401	
T	TRANSMEM	402	404	LOOP BETWEEN TWO HELICES.
T	TRANSMEM	405	416	
T	TRANSMEM	417	421	LOOP BETWEEN TWO HELICES.
T	TRANSMEM	422	438	
T	DOMAIN	439	473	CYTOPLASMIC.
T	SITE	148	148	MAY BE INVOLVED IN GATING.
T	BINDING	107	107	CHLORIDE.
T	BINDING	356	356	CHLORIDE (VIA AMIDE NITROGEN).
T	BINDING	357	357	CHLORIDE (VIA AMIDE NITROGEN).
T	BINDING	445	445	CHLORIDE.
T	CONFLICT	32	32	P -> Q (IN REF. 1).
T	HELIX	13	25	
T	TURN	26	27	
T	HELIX	33	64	
T	TURN	65	70	
T	HELIX	75	100	
T	TURN	102	103	
T	HELIX	109	115	
T	TURN	116	117	
T	HELIX	124	141	
T	TURN	142	142	
T	STRAND	146	146	
T	HELIX	148	165	
T	TURN	166	167	
T	TURN	170	171	
T	HELIX	172	190	
T	TURN	191	191	
T	HELIX	193	199	
T	TURN	200	202	
T	HELIX	215	232	
T	HELIX	249	251	
T	HELIX	252	275	
T	TURN	276	286	
T	TURN	288	289	
T	HELIX	290	305	
T	TURN	306	307	
T	HELIX	310	312	
T	TURN	317	319	
T	HELIX	320	324	
T	TURN	325	326	
T	HELIX	330	350	
T	TURN	351	351	
T	STRAND	355	355	
T	TURN	357	358	
T	HELIX	359	378	
T	TURN	380	381	
T	HELIX	386	392	
T	TURN	393	394	
T	HELIX	396	401	
T	TURN	402	402	
T	HELIX	405	416	
T	HELIX	419	421	
T	HELIX	422	434	
T	TURN	435	440	
T	HELIX	444	457	
T	TURN	458	459	

Query Match 83.3%; Score 35; DB 1; Length 473;
Best Local Similarity 44.4%; Pred. No. 46;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VVLGVVFGV 9
              ::||::||:
Db        256 LILGIIFGI 264
```

Search completed: May 17, 2004, 12:57:00
Job time : 6.96774 secs

GenCore version 5.1.6

M protein - protein search, using sw model

un on: May 17, 2004, 12:47:22 ; Search time 28.7419 Seconds
(without alignments)
98.799 Million cell updates/sec

itle: US-09-458-299A-4239
erfect score: 42
equence: 1 VVLGVVFGV 9

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 1017041 seqs, 315518202 residues

otal number of hits satisfying chosen parameters: 1017041

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	41	97.6	2972	5	P90891	P90891 caenorhabdi
2	38	90.5	128	17	Q8U416	Q8u416 pyrococcus
3	38	90.5	258	16	Q897U8	Q897u8 clostridium
4	38	90.5	408	16	Q8RF88	Q8rf88 fusobacteri
5	37	88.1	172	16	Q8FQD4	Q8fqd4 corynebacte
6	37	88.1	331	17	Q8U3G5	Q8u3g5 pyrococcus
7	37	88.1	342	7	O19479	O19479 ginglymosto
8	37	88.1	377	17	O30022	O30022 archaeoglob
9	37	88.1	388	16	Q89KE8	Q89ke8 bradyrhizob
10	37	88.1	413	5	Q8ICC1	Q8icc1 plasmodium
11	37	88.1	521	2	Q9RMF9	Q9rmf9 streptomyce
12	36	85.7	57	16	Q9PPD3	Q9ppd3 campylobact
13	36	85.7	65	16	Q8DMT8	Q8dmt8 synechococc
14	36	85.7	93	16	O24860	O24860 helicobacte
15	36	85.7	93	16	Q9ZN46	Q9zn46 helicobacte
16	36	85.7	145	17	Q8TMX8	Q8tmx8 methanosarc
17	36	85.7	220	16	Q89BN0	Q89bn0 bradyrhizob
18	36	85.7	297	16	Q82KH0	Q82kh0 streptomyce
19	36	85.7	344	2	Q8KN75	Q8kn75 pseudomonas
20	36	85.7	385	16	Q9PFR8	Q9pfr8 xylella fas
21	36	85.7	385	16	Q87B96	Q87b96 xylella fas
22	36	85.7	421	2	Q9ZNF6	Q9znf6 bacillus li
23	36	85.7	429	16	Q97FG7	Q97fg7 clostridium
24	36	85.7	449	16	Q8RCG1	Q8rcg1 thermoanaer
25	36	85.7	451	16	P73745	P73745 synechocyst
26	36	85.7	463	16	Q88CP9	Q88cp9 pseudomonas
27	36	85.7	799	16	Q89HX5	Q89hx5 bradyrhizob
28	36	85.7	822	16	Q8YYK7	Q8yyk7 anabaena sp
29	35	83.3	35	17	Q9HNF2	Q9hnf2 halobacteri
30	35	83.3	80	9	Q9AZK2	Q9azk2 bacteriopha
31	35	83.3	80	16	Q9CJF2	Q9cjf2 lactococcus
32	35	83.3	149	10	O49442	O49442 arabidopsis
33	35	83.3	159	10	O49443	O49443 arabidopsis
34	35	83.3	165	17	Q8THE9	Q8the9 methanosarc
35	35	83.3	200	16	Q8FZ71	Q8fz71 brucella su
36	35	83.3	202	10	Q8LNV2	Q8lnv2 oryza sativ
37	35	83.3	204	16	Q8YIP0	Q8yip0 brucella me
38	35	83.3	219	16	Q9PMG9	Q9pmg9 campylobact
39	35	83.3	234	16	Q9EWV1	Q9ewv1 streptomyce
40	35	83.3	284	2	Q8G9N0	Q8g9n0 rhodococcus
41	35	83.3	308	16	Q8NP34	Q8np34 corynebacte
42	35	83.3	320	16	Q83Q35	Q83q35 shigella fl
43	35	83.3	321	16	Q8XAJ0	Q8xaj0 escherichia
44	35	83.3	321	16	Q8FDE1	Q8fde1 escherichia
45	35	83.3	321	16	Q7UBI0	Q7ubi0 shigella fl

ALIGNMENTS

RESULT 1
P90891
ID P90891 PRELIMINARY; PRT; 2972 AA.
AC P90891;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F55H12.3 protein.
GN F55H12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81091; CAB03143.2; -.
DR PIR; T22759; T22759.
DR HSSP; P00740; 1EDM.
DR WormPep; F55H12.3; CE25008.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003410; Hyalin.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001304; Lectin_C.

```
)R   InterPro; IPR000436; Sushi_SCR_CCP.
)R   Pfam; PF00008; EGF; 7.
)R   Pfam; PF02494; HYR; 1.
)R   Pfam; PF00057; ldl_recept_a; 1.
)R   Pfam; PF00059; lectin_c; 1.
)R   Pfam; PF00084; sushi; 3.
)R   PRINTS; PR00011; EGFLAMININ.
)R   SMART; SM00032; CCP; 3.
)R   SMART; SM00034; CLECT; 1.
)R   SMART; SM00181; EGF; 13.
)R   SMART; SM00192; LDLa; 1.
)R   PROSITE; PS00010; ASX_HYDROXYL; 1.
)R   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
)R   PROSITE; PS00022; EGF_1; 8.
)R   PROSITE; PS01186; EGF_2; 4.
)R   PROSITE; PS01187; EGF_CA; 1.
)R   PROSITE; PS50068; LDLRA_2; 1.
CW   EGF-like domain.
SQ   SEQUENCE   2972 AA;  329457 MW;  87D7BA80562F4C74 CRC64;

 Query Match             97.6%;  Score 41;  DB 5;  Length 2972;
 Best Local Similarity   88.9%;  Pred. No. 1.2e+02;
 Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

)y          1 VVLGVVFGV 9
              :||||||||
)b       2799 IVLGVVFGV 2807


ESULT 2
8U416
D   Q8U416      PRELIMINARY;      PRT;   128 AA.
C   Q8U416;
T   01-JUN-2002 (TrEMBLrel. 21, Created)
T   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
T   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
E   Hypothetical protein PF0281.
N   PF0281.
S   Pyrococcus furiosus.
C   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
C   Pyrococcus.
X   NCBI_TaxID=2261;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
A   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
T   "The complete sequence of the Pyrococcus furiosus genome.";
L   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AE010152; AAL80405.1; -.
W   Hypothetical protein; Complete proteome.
Q   SEQUENCE   128 AA;  13112 MW;  91D0FF21623573DF CRC64;

 Query Match             90.5%;  Score 38;  DB 17;  Length 128;
 Best Local Similarity   55.6%;  Pred. No. 22;
 Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

y          1 VVLGVVFGV 9
             ::||:|||:
b        114 IILGIVFGI 122


ESULT 3
897U8
D   Q897U8      PRELIMINARY;      PRT;   258 AA.
C   Q897U8;
T   01-JUN-2003 (TrEMBLrel. 24, Created)
T   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E   Lantibiotic transport-associated permease spaG/mutG.
N   SPAG OR CTC00627.
S   Clostridium tetani.
C   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
C   Clostridium.
```

```
OX   NCBI_TaxID=1513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Massachusetts / E88;
RX   MEDLINE=22457253; PubMed=12552129;
RA   Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,
RA   Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA   Gottschalk G.;
RT   "The genome sequence of Clostridium tetani, the causative agent of
RT   tetanus disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR   EMBL; AE015938; AAO35238.1; -.
KW   Complete proteome.
SQ   SEQUENCE   258 AA;  29217 MW;  51B0A814C1C64BCE CRC64;

  Query Match             90.5%;  Score 38;  DB 16;  Length 258;
  Best Local Similarity   55.6%;  Pred. No. 43;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::||:|||:
Db        239 IILGIVFGI 247


RESULT 4
Q8RF88
ID   Q8RF88      PRELIMINARY;      PRT;   408 AA.
AC   Q8RF88;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   ABC transporter permease protein.
GN   FN0828.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterales; Fusobacteriaceae;
OC   Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010592; AAL95024.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   InterPro; IPR003838; DUF214.
DR   Pfam; PF02687; FtsX; 1.
KW   Complete proteome.
SQ   SEQUENCE   408 AA;  45050 MW;  926EC1271F3EC494 CRC64;

  Query Match             90.5%;  Score 38;  DB 16;  Length 408;
  Best Local Similarity   77.8%;  Pred. No. 67;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ||:||:|||
Db        382 VVVGVIFGV 390


RESULT 5
Q8FQD4
ID   Q8FQD4      PRELIMINARY;      PRT;   172 AA.
AC   Q8FQD4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Conserved hypothetical protein.
```

```
:N   CE1185.
:S   Corynebacterium efficiens.
:C   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
:C   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
:X   NCBI_TaxID=152794;
:N   [1]
:P   SEQUENCE FROM N.A.
:C   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
:A   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
:A   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
:A   Usuda Y., Sugimoto S.;
:T   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
:L   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
:R   EMBL; AP005218; BAC17995.1; -.
:W   Hypothetical protein; Complete proteome.
:Q   SEQUENCE   172 AA;  18639 MW;  4C39A6DA55C3CA7D CRC64;

 Query Match             88.1%;  Score 37;  DB 16;  Length 172;
 Best Local Similarity   77.8%;  Pred. No. 45;
 Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Y          1 VVLGVVFGV 9
             :|:||||||
b        109 LVMGVVFGV 117


ESULT 6
8U3G5
D   Q8U3G5      PRELIMINARY;      PRT;   331 AA.
C   Q8U3G5;
T   01-JUN-2002 (TrEMBLrel. 21, Created)
T   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
T   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E   Putative iron ABC transporter.
N   PF0503.
S   Pyrococcus furiosus.
C   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
C   Pyrococcus.
X   NCBI_TaxID=2261;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
A   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
T   "The complete sequence of the Pyrococcus furiosus genome.";
L   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AE010173; AAL80627.1; -.
R   GO; GO:0016020; C:membrane; IEA.
R   GO; GO:0005311; F:sodium:dicarboxylate/tricarboxylate symport. . .; IEA.
R   GO; GO:0005215; F:transporter activity; IEA.
R   GO; GO:0006835; P:dicarboxylic acid transport; IEA.
R   GO; GO:0006810; P:transport; IEA.
R   InterPro; IPR000522; FecD.
R   InterPro; IPR001991; Na/diCO_symport.
R   Pfam; PF01032; FecCD; 1.
R   PRINTS; PR00173; EDTRNSPORT.
R   ProDom; PD001557; FecCD; 1.
W   Complete proteome; Hypothetical protein.
Q   SEQUENCE   331 AA;  35473 MW;  8D2545E94BD70CA1 CRC64;

 Query Match             88.1%;  Score 37;  DB 17;  Length 331;
 Best Local Similarity   87.5%;  Pred. No. 84;
 Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

/          1 VVLGVVFG 8
             |:||||||
)         16 VILGVVFG 23


ESULT 7
.9479
)   O19479      PRELIMINARY;      PRT;   342 AA.
:   O19479;
T   01-JAN-1998 (TrEMBLrel. 05, Created)
T   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
T   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E   MHC class I protein.
N   GICI-11.
S   Ginglymostoma cirratum (Nurse shark).
C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
C   Elasmobranchii; Galeomorphii; Galeoidea; Orectolobiformes;
C   Ginglymostomatidae; Ginglymostoma.
X   NCBI_TaxID=7801;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=98209780; PubMed=9550410;
A   Bartl S., Baish M.A., Flajnik M.F., Ohta Y.;
T   "Identification of class I genes in cartilaginous fish, the most
T   ancient group of vertebrates displaying an adaptive immune response.";
L   J. Immunol. 159:6097-6104(1997).
C   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
C       IMMUNE SYSTEM (BY SIMILARITY).
C   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
C       MICROGLOBULIN) (BY SIMILARITY).
R   EMBL; AF028557; AAC60347.1; -.
R   GO; GO:0016021; C:integral to membrane; IEA.
R   GO; GO:0006955; P:immune response; IEA.
R   InterPro; IPR007110; Ig-like.
R   InterPro; IPR003597; Ig_c1.
R   InterPro; IPR001039; MHC_I.
R   Pfam; PF00047; ig; 1.
R   Pfam; PF00129; MHC_I; 1.
R   PRINTS; PR01638; MHCCLASSI.
R   ProDom; PD000050; MHC_I; 1.
R   SMART; SM00407; IGc1; 1.
R   PROSITE; PS50835; IG_LIKE; 1.
K   Glycoprotein; Transmembrane.
Q   SEQUENCE   342 AA;  38926 MW;  DC847FB53DBF9C01 CRC64;

  Query Match             88.1%;  Score 37;  DB 7;  Length 342;
  Best Local Similarity   77.8%;  Pred. No. 87;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         1 VVLGVVFGV 9
             | ||:||||
Db       290 VTLGIVFGV 298


RESULT 8
O30022
ID   O30022      PRELIMINARY;      PRT;   377 AA.
AC   O30022;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   NA+/H+ antiporter (NAPA-1).
GN   AF0217.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
```

```
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001091; AAB91016.1; -.
DR   PIR; A69277; A69277.
DR   TIGR; AF0217; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0008324; F:cation transporter activity; IEA.
DR   GO; GO:0015299; F:solute:hydrogen antiporter activity; IEA.
DR   GO; GO:0006812; P:cation transport; IEA.
DR   GO; GO:0006885; P:regulation of pH; IEA.
DR   InterPro; IPR004771; K_eff.
DR   InterPro; IPR006153; Na_H_porter.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
DR   TIGRFAMs; TIGR00932; 2a37; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   377 AA;  41026 MW;  7348380033E12F8C CRC64;

  Query Match             88.1%;  Score 37;  DB 17;  Length 377;
  Best Local Similarity   66.7%;  Pred. No. 95;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              :|||::|||
Db         36 IVLGMIFGV 44


RESULT 9
Q89KE8
ID   Q89KE8      PRELIMINARY;      PRT;   388 AA.
AC   Q89KE8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Blr4959 protein.
GN   BLR4959.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005953; BAC50224.1; -.
DR   InterPro; IPR002549; UPF0118.
DR   Pfam; PF01594; UPF0118; 1.
KW   Complete proteome.
SQ   SEQUENCE   388 AA;  40604 MW;  BE2711E3D9DF177E CRC64;

  Query Match             88.1%;  Score 37;  DB 16;  Length 388;
  Best Local Similarity   66.7%;  Pred. No. 98;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              ::||||||:
Db        348 ILLGVVFGI 356


RESULT 10
Q8ICC1
ID   Q8ICC1      PRELIMINARY;      PRT;   413 AA.
AC   Q8ICC1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   MAL6P1.177.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844505; CAD50521.1; -.
DR   InterPro; IPR000326; PA_PTPase.
KW   Hypothetical protein.
SQ   SEQUENCE   413 AA;  48546 MW;  B3B8C8D2E4E1685F CRC64;

  Query Match             88.1%;  Score 37;  DB 5;  Length 413;
  Best Local Similarity   66.7%;  Pred. No. 1e+02;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFGV 9
              |:||::|||
Db        270 VLLGIIFGV 278


RESULT 11
Q9RMF9
ID   Q9RMF9      PRELIMINARY;      PRT;   521 AA.
AC   Q9RMF9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SgcB (Transmembrane efflux protein).
GN   SGCB.
OS   Streptomyces globisporus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1908;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20106985; PubMed=10639366;
RA   Wen L., Shen B.;
RT   "Genes for production of the enediyne antitumor antibiotic C-1027 in
RT   Streptomyces globisporus are clustered with the cagA gene that encodes
RT   the C-1027 apoprotein.";
RL   Antimicrob. Agents Chemother. 44:382-392(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-1027;
RX   MEDLINE=22171413; PubMed=12183628;
RA   Liu W., Christenson S.D., Standage S., Shen B.;
RT   "Biosynthesis of the enediyne antitumor antibiotic C-1027.";
RL   Science 297:1170-1173(2002).
DR   EMBL; AF201913; AAF13999.1; -.
DR   EMBL; AY048670; AAL06672.1; -.
DR   InterPro; IPR007114; MFS.
DR   PROSITE; PS50850; MFS; 1.
SQ   SEQUENCE   521 AA;  52953 MW;  77B604EBF69856C9 CRC64;

  Query Match             88.1%;  Score 37;  DB 2;  Length 521;
  Best Local Similarity   87.5%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VVLGVVFG 8
              |||||:||
Db        239 VVLGVIFG 246


RESULT 12
Q9PPD3
ID   Q9PPD3      PRELIMINARY;      PRT;    57 AA.
AC   Q9PPD3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
T   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E   Small hydrophobic protein.
N   CJ0786.
S   Campylobacter jejuni.
C   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
C   Campylobacteraceae; Campylobacter.
X   NCBI_TaxID=197;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=NCTC 11168;
K   MEDLINE=20150912; PubMed=10688204;
A   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
A   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
A   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
A   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
A   Whitehead S., Barrell B.G.;
T   "The genome sequence of the food-borne pathogen Campylobacter jejuni
T   reveals hypervariable sequences.";
L   Nature 403:665-668(2000).
R   EMBL; AL139076; CAB73051.1; -.
R   PIR; B81350; B81350.
W   Complete proteome.
Q   SEQUENCE   57 AA;  6821 MW;  4935834C4E384B62 CRC64;

 Query Match             85.7%;  Score 36;  DB 16;  Length 57;
 Best Local Similarity   44.4%;  Pred. No. 25;
 Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

          1 VVLGVVFGV 9
            :::||:||:
          9 IIIGVIFGI 17


SULT 13
3DMT8
)   Q8DMT8      PRELIMINARY;      PRT;    65 AA.
:   Q8DMT8;
:   01-MAR-2003 (TrEMBLrel. 23, Created)
:   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
:   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
3   Tsl0023 protein.
N   TSL0023.
3   Synechococcus elongatus (Thermosynechococcus elongatus).
2   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
(   NCBI_TaxID=32046;
N   [1]
P   SEQUENCE FROM N.A.
:   STRAIN=BP-1;
(   MEDLINE=22225144; PubMed=12240834;
A   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
A   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
A   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
A   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
T   "Complete genome structure of the thermophilic cyanobacterium
T   Thermosynechococcus elongatus BP-1.";
L   DNA Res. 9:123-130(2002).
R   EMBL; AP005369; BAC07576.1; -.
W   Complete proteome.
Q   SEQUENCE   65 AA;  6790 MW;  F6B8AA287DC930AC CRC64;

 Query Match             85.7%;  Score 36;  DB 16;  Length 65;
 Best Local Similarity   55.6%;  Pred. No. 28;
 Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

          1 VVLGVVFGV 9
            ::|||:||:
         13 MILGVIFGI 21


SULT 14
4860
)   O24860      PRELIMINARY;      PRT;    93 AA.
:   O24860;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein HP0015.
GN   HP0015.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000524; AAD07091.1; -.
DR   PIR; G64521; G64521.
DR   TIGR; HP0015; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   93 AA;  10526 MW;  5E13E652C402A22F CRC64;

  Query Match             85.7%;  Score 36;  DB 16;  Length 93;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VLGVVFGV 9
              :||||||:
Db         69 ILGVVFGI 76


RESULT 15
Q9ZN46
ID   Q9ZN46      PRELIMINARY;      PRT;    93 AA.
AC   Q9ZN46;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative.
GN   JHP0013.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001441; AAD05595.1; -.
DR   PIR; C71984; C71984.
KW   Complete proteome.
SQ   SEQUENCE   93 AA;  10512 MW;  5E017652C402A22F CRC64;

  Query Match             85.7%;  Score 36;  DB 16;  Length 93;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
y          2 VLGVVFGV 9
             :||||||:
b         69 ILGVVFGI 76
```

earch completed: May 17, 2004, 12:56:27
ob time : 30.7419 secs

GenCore version 5.1.6

M protein - protein search, using sw model

un on: May 17, 2004, 13:31:59 ; Search time 54 Seconds
(without alignments)
68.021 Million cell updates/sec

itle: US-09-458-299A-4226
erfect score: 55
equence: 1 XKXVWANTLKAAX 13

coring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

earched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	55	100.0	13	2	AAR75944	Aar75944 T helper
2	55	100.0	13	4	AAJ04120	Aaj04120 Pan-DR bi
3	41	74.5	13	3	AAY99331	Aay99331 HLA class
4	41	74.5	13	4	AAB99718	Aab99718 Pan-DR-bi
5	41	74.5	13	4	AAB99719	Aab99719 Pan-DR-bi
6	41	74.5	13	4	AAB99720	Aab99720 Pan-DR-bi
7	40	72.7	425	4	AAB76652	Aab76652 Corynebac
8	40	72.7	425	4	AAB76651	Aab76651 Corynebac
9	40	72.7	425	4	AAB80185	Aab80185 Corynebac
10	40	72.7	537	2	AAR80504	Aar80504 S. livida
11	40	72.7	537	2	AAW87796	Aaw87796 Tripeptid
12	40	72.7	537	3	AAB19228	Aab19228 A tripept
13	40	72.7	541	5	ABB07853	Abb07853 C. glutam
14	40	72.7	543	4	AAG91463	Aag91463 C glutami
15	39	70.9	361	3	AAG10400	Aag10400 Arabidops
16	39	70.9	390	3	AAG10399	Aag10399 Arabidops
17	39	70.9	423	3	AAG10398	Aag10398 Arabidops
18	38	69.1	55	3	AAY71540	Aay71540 Wheat POP
19	38	69.1	293	7	ADC39192	Adc39192 Novel hum
20	38	69.1	294	3	AAY71536	Aay71536 Wheat POP
21	38	69.1	402	3	AAB53142	Aab53142 Macaca mu
22	38	69.1	439	4	ABG08417	Abg08417 Novel hum
23	38	69.1	460	3	AAB58987	Aab58987 Breast an
24	38	69.1	460	5	ABP41511	Abp41511 Human ova
25	38	69.1	498	2	AAY34725	Aay34725 Chlamydia
26	38	69.1	585	6	ADA34846	Ada34846 Acinetoba
27	37	67.3	13	3	AAY99329	Aay99329 HLA class
28	37	67.3	125	3	AAB33180	Aab33180 Eucalyptu
29	37	67.3	351	3	AAG29403	Aag29403 Arabidops
30	37	67.3	358	3	AAG29402	Aag29402 Arabidops
31	37	67.3	361	3	AAG14378	Aag14378 Arabidops
32	37	67.3	368	5	AAU98468	Aau98468 Transcrip
33	37	67.3	390	3	AAG14377	Aag14377 Arabidops
34	37	67.3	398	3	AAG29401	Aag29401 Arabidops
35	37	67.3	398	7	ADD30005	Add30005 Plant yie
36	37	67.3	423	3	AAG14376	Aag14376 Arabidops
37	37	67.3	423	7	ADD31013	Add31013 Plant yie
38	37	67.3	444	4	ABB68897	Abb68897 Drosophil
39	37	67.3	807	5	ABB93567	Abb93567 Herbicida
40	37	67.3	1058	4	ABB62096	Abb62096 Drosophil
41	37	67.3	1096	6	ABU43426	Abu43426 Protein e
42	36	65.5	13	3	AAY99110	Aay99110 HLA class
43	36	65.5	13	3	AAY99115	Aay99115 HLA class
44	36	65.5	13	4	AAB99717	Aab99717 Pan-DR-bi
45	36	65.5	13	4	AAB99716	Aab99716 Pan-DR-bi

ALIGNMENTS

RESULT 1
AAR75944
ID AAR75944 standard; peptide; 13 AA.
XX
AC AAR75944;
XX
DT 19-MAR-1996 (first entry)
XX
DE T helper epitope.
XX
KW MAGE-3; melanoma antigen; vaccine; immune response; immunogenic peptide;
KW cytotoxic T lymphocyte response; CTL; melanoma; breast cancer; antibody.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 3
FT /note= "unidentified in the specification"
XX
PN WO9519783-A1.
XX
PD 27-JUL-1995.
XX
PF 25-JAN-1995; 95WO-US001000.
XX
PR 25-JAN-1994; 94US-00186266.
XX
PA (CYTE-) CYTEL CORP.
XX
PI Kubo RT, Grey HM, Sette A, Celis E;
XX
DR WPI; 1995-269270/35.
XX
PT Immunogenic peptide(s) that induce immune response to cancer cells - that
PT express a MAGE-3 protein peptide epitope used in vaccines or adoptive
PT immuno:therapy to induce cytotoxic T lymphocytes.
XX
PS Claim 21; Page 37; 44pp; English.
XX
CC AAR75942 is derived from the sequence of the melanoma antigen (MAGE-3)
CC protein and can be used to elicit a primary cytotoxic T lymphocyte
CC response against cells expressing MAGE-3. Synthetic peptides AAR75945-53
CC can be used therapeutically to elicit CTL responses to melanoma, breast,
CC colon, prostate, or other cells which express proteins with this epitope.
CC The peptides have specific HLA-A1 binding capacity. The peptides can be
CC also used in vaccines, esp. combined with peptides such as AAR75943-44,
CC which are T-helper epitopes
XX

Q Sequence 13 AA;

Query Match 100.0%; Score 55; DB 2; Length 13;
Best Local Similarity 84.6%; Pred. No. 0.0055;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

y 1 XKXVWANTLKAAX 13
:||||||||||:
b 1 AKXVWANTLKAAA 13

ESULT 2
AJ04120
D AAJ04120 standard; peptide; 13 AA.
X
C AAJ04120;
X
T 02-JUL-2001 (first entry)
X
E Pan-DR binding epitope.
X
W Hepatitis C virus; HCV; epitope; vaccine; immunogen; HLA-binding motif;
W antiviral.
X
S Synthetic.
X
N WO200121189-A1.
X
D 29-MAR-2001.
X
F 19-JUL-2000; 2000WO-US019774.
X
R 19-JUL-1999; 99US-00357737.
X
A (EPIM-) EPIMMUNE INC.
X
I Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
I Baker DM, Celis E, Kubo RT, Grey HM;
X
R WPI; 2001-308046/32.
X
T A new composition useful as a vaccines against hepatitis C virus.
X
S Disclosure; Page 53; 214pp; English.
X
C The present invention describes a composition comprising a prepared
C hepatitis C virus (HCV) epitope such as those given in AAJ00010-AAJ04121.
C These are derived from HCV HLA-binding motifs. They are useful in
C vaccines for the prevention and treatment of HCV infection in humans. The
C present sequence is an epitope used in the disclosure of the invention
X
Q Sequence 13 AA;

Query Match 100.0%; Score 55; DB 4; Length 13;
Best Local Similarity 84.6%; Pred. No. 0.0055;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

y 1 XKXVWANTLKAAX 13
:||||||||||:
b 1 AKXVWANTLKAAA 13

ESULT 3
AY99331
D AAY99331 standard; peptide; 13 AA.
X
C AAY99331;
X
T 07-AUG-2000 (first entry)
X
E HLA class II binding antigen epitope peptide #520.
X
W Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
KW immune response; chronic viral disease; cancer; autoimmune disease;
KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
KW allograft rejection; allergy; lyme disease; hepatitis; prostate cancer;
KW glomerulonephritis; food hypersensitivity; malaria.
XX
OS Unidentified.
XX
PN WO9961916-A1.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US012066.
XX
PR 29-MAY-1998; 98US-0087192P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Southwood S, Sidney J;
XX
DR WPI; 2000-097143/08.
XX
PT New compositions containing immunogenic peptide epitopes for various HLA
PT class II DR molecules useful for inducing helper T cell response.
XX
PS Claim 1; Page 48; 60pp; English.
XX
CC The present invention relates to a new pharmaceutical composition
CC comprising a unit dose form of a peptide, or analogue, comprising an
CC epitope selected from those represented by peptides AAY98812-Y99339 which
CC are derived from various antigens for various human leucocyte antigen
CC class DR molecules, representative of the world wide population. The
CC peptide/analogue binds to an HLA class II molecule at an IC-50 of less
CC than or equal to 1,000 nM. The pharmaceutical can be used to induce a
CC helper T cell response. The pharmaceutical focuses the immune response
CC towards selected determinants and could therefore be used in cases of
CC chronic viral diseases and cancer. Examples of diseases that can be
CC treated using the peptide containing pharmaceutical include autoimmune
CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
CC gravis), allograft rejection, allergies, lyme disease, hepatitis, post-
CC streptococcal endocarditis or glomerulonephritis and food
CC hypersensitivities. The peptide epitopes can be used to enhance immune
CC responses against other immunogens administered with the peptides.
CC Diseases which can be treated using immunogenic mixtures include prostate
CC cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma, cervical
CC carcinoma, lymphoma, and condyloma acuminatum. The peptides may also be
CC used to make monoclonal antibodies useful as potential diagnostic or
CC therapeutic agents. The peptides may also be useful as diagnostic
CC reagents, for example, to determine the susceptibility of an individual
CC to a treatment regimen. Also, the peptides may be used to predict which
CC individuals will be at substantial risk of developing chronic infection.
CC The selection of appropriate T and B cell epitopes should allow the
CC development of epitope based vaccines particularly towards conserved
CC epitopes of pathogens which are characterized by high sequence
CC variability such as HIV, HCV and Malaria
XX
SQ Sequence 13 AA;

Query Match 74.5%; Score 41; DB 3; Length 13;
Best Local Similarity 69.2%; Pred. No. 1.6;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 XKXVWANTLKAAX 13
:|:| |||||||:
Db 1 AKFVKANTLKAAA 13

RESULT 4
AAB99718
ID AAB99718 standard; peptide; 13 AA.
XX
AC AAB99718;
XX
DT 06-SEP-2001 (first entry)

Pan-DR-binding peptide (PADRE) SEQ ID NO:39.

Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL; cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu; MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator; immunotherapy; immune response.

Homo sapiens.
Synthetic.

WO200141741-A1.

14-JUN-2001.

13-DEC-2000; 2000WO-US034318.

13-DEC-1999; 99US-0170448P.
05-APR-2000; 2000US-00543608.
30-MAY-2000; 2000US-00583200.

(EPIM-) EPIMMUNE INC.

Fikes J, Sette A, Sidney J, Southwood S, Celis E, Keogh E; Chesnut R;

WPI; 2001-381489/40.

Compositions for use in a vaccine for treating, e.g., breast, lung and colon cancer comprises at least one peptide that comprises an isolated epitope of a tumor-associated antigen.

Example 7; Page 48; 86pp; English.

The present invention describes a composition (I) comprising at least one peptide that comprises an isolated, prepared epitope consisting of a sequence selected from 25 short amino acid sequences given in AAB99680 to AAB99704. Also described are: (1) a composition (II) comprising one or more peptides, and further comprising at least two epitopes selected from the 25 short amino acid sequences (as above), where each of the one or more peptides comprise less than 50 contiguous amino acids that have 100% identity with a native peptide sequence; and (2) a vaccine composition (III) comprising an epitope selected from the 25 short amino acid sequences (as above) and a pharmaceutical excipient. (I) has cytostatic and immunomodulatory activities and can be used in vaccine production and immunotherapy. The peptide epitope compositions (I)-(II) are useful for monitoring an immune response to a tumour associated antigen or when one or more peptides are combined to create a vaccine (III) that stimulates the cellular arm of the immune system. In particular, the vaccine mediates immune responses against tumours in individuals who bear an allele of the human leukocyte antigen (HLA)-A2 supertype and improve the standard of care for patients being treated for breast, colon, or lung cancer. The present sequence represents a pan-DR-binding peptide (PADRE) sequence, which is used in an example from the present invention

Sequence 13 AA;

```
Query Match             74.5%;  Score 41;  DB 4;  Length 13;
Best Local Similarity   69.2%;  Pred. No. 1.6;
Matches    9;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

          1 XKXVWANTLKAAX 13
            :|:| |||||||:
          1 AKFVAANTLKAAA 13
```

```
SULT 5
B99719
   AAB99719 standard; peptide; 13 AA.

   AAB99719;

   06-SEP-2001  (first entry)
```

```
XX
DE   Pan-DR-binding peptide (PADRE) SEQ ID NO:40.
XX
KW   Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL;
KW   cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu;
KW   MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator;
KW   immunotherapy; immune response.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   3
FT                   /note= "cyclohexylalanine"
XX
PN   WO200141741-A1.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-US034318.
XX
PR   13-DEC-1999;   99US-0170448P.
PR   05-APR-2000; 2000US-00543608.
PR   30-MAY-2000; 2000US-00583200.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes J,  Sette A,  Sidney J,  Southwood S,  Celis E,  Keogh E;
PI   Chesnut R;
XX
DR   WPI; 2001-381489/40.
XX
PT   Compositions for use in a vaccine for treating, e.g., breast, lung and
PT   colon cancer comprises at least one peptide that comprises an isolated
PT   epitope of a tumor-associated antigen.
XX
PS   Example 7; Page 48; 86pp; English.
XX
CC   The present invention describes a composition (I) comprising at least one
CC   peptide that comprises an isolated, prepared epitope consisting of a
CC   sequence selected from 25 short amino acid sequences given in AAB99680 to
CC   AAB99704. Also described are: (1) a composition (II) comprising one or
CC   more peptides, and further comprising at least two epitopes selected from
CC   the 25 short amino acid sequences (as above), where each of the one or
CC   more peptides comprise less than 50 contiguous amino acids that have 100%
CC   identity with a native peptide sequence; and (2) a vaccine composition
CC   (III) comprising an epitope selected from the 25 short amino acid
CC   sequences (as above) and a pharmaceutical excipient. (I) has cytostatic
CC   and immunomodulatory activities and can be used in vaccine production and
CC   immunotherapy. The peptide epitope compositions (I)-(II) are useful for
CC   monitoring an immune response to a tumour associated antigen or when one
CC   or more peptides are combined to create a vaccine (III) that stimulates
CC   the cellular arm of the immune system. In particular, the vaccine
CC   mediates immune responses against tumours in individuals who bear an
CC   allele of the human leukocyte antigen (HLA)-A2 supertype and improve the
CC   standard of care for patients being treated for breast, colon, or lung
CC   cancer. The present sequence represents a pan-DR-binding peptide (PADRE)
CC   sequence, which is used in an example from the present invention
XX
SQ   Sequence 13 AA;

  Query Match             74.5%;  Score 41;  DB 4;  Length 13;
  Best Local Similarity   76.9%;  Pred. No. 1.6;
  Matches   10;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :||| |||||||:
Db          1 AKXVAANTLKAAA 13
```

```
RESULT 6
AAB99720
ID   AAB99720 standard; peptide; 13 AA.
```

```
X
C   AAB99720;
X
T   06-SEP-2001  (first entry)
X
E   Pan-DR-binding peptide (PADRE) SEQ ID NO:41.
X
W   Human leukocyte antigen A2 binding peptide; HLA class I A2; CTL;
W   cytotoxic T-cell lymphocyte; tumour associated antigen; CEA; HER2/neu;
W   MAGE2; MAGE3; p53; vaccine; cancer; cytostatic; immunomodulator;
W   immunotherapy; immune response.
X
S   Homo sapiens.
S   Synthetic.
X
N   WO200141741-A1.
X
D   14-JUN-2001.
X
F   13-DEC-2000; 2000WO-US034318.
X
R   13-DEC-1999;   99US-0170448P.
R   05-APR-2000; 2000US-00543608.
R   30-MAY-2000; 2000US-00583200.
X
A   (EPIM-) EPIMMUNE INC.
X
I   Fikes J,  Sette A,  Sidney J,  Southwood S,  Celis E,  Keogh E;
I   Chesnut R;
X
R   WPI; 2001-381489/40.
X
T   Compositions for use in a vaccine for treating, e.g., breast, lung and
T   colon cancer comprises at least one peptide that comprises an isolated
T   epitope of a tumor-associated antigen.
X
S   Example 7; Page 48; 86pp; English.
X
C   The present invention describes a composition (I) comprising at least one
C   peptide that comprises an isolated, prepared epitope consisting of a
C   sequence selected from 25 short amino acid sequences given in AAB99680 to
C   AAB99704. Also described are: (1) a composition (II) comprising one or
C   more peptides, and further comprising at least two epitopes selected from
C   the 25 short amino acid sequences (as above), where each of the one or
C   more peptides comprise less than 50 contiguous amino acids that have 100%
C   identity with a native peptide sequence; and (2) a vaccine composition
C   (III) comprising an epitope selected from the 25 short amino acid
C   sequences (as above) and a pharmaceutical excipient. (I) has cytostatic
C   and immunomodulatory activities and can be used in vaccine production and
C   immunotherapy. The peptide epitope compositions (I)-(II) are useful for
C   monitoring an immune response to a tumour associated antigen or when one
C   or more peptides are combined to create a vaccine (III) that stimulates
C   the cellular arm of the immune system. In particular, the vaccine
C   mediates immune responses against tumours in individuals who bear an
C   allele of the human leukocyte antigen (HLA)-A2 supertype and improve the
C   standard of care for patients being treated for breast, colon, or lung
C   cancer. The present sequence represents a pan-DR-binding peptide (PADRE)
C   sequence, which is used in an example from the present invention
X
Q   Sequence 13 AA;

 Query Match             74.5%;  Score 41;  DB 4;  Length 13;
 Best Local Similarity   69.2%;  Pred. No. 1.6;
 Matches    9;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

y          1 XKXVWANTLKAAX 13
             :|:| |||||||:
b          1 AKYVAANTLKAAA 13


ESULT 7
AB76652
D   AAB76652 standard; protein; 425 AA.
XX
AC   AAB76652;
XX
DT   11-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum MCT protein SEQ ID NO:286.
XX
KW   Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW   membrane construction and membrane transport protein; petroleum spill;
KW   hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW   identification; microorganism; fine chemical production; transformation;
KW   genome mapping; genetic engineering.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100805-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB000926.
XX
PR   25-JUN-1999;   99US-0141031P.
PR   08-JUL-1999;   99DE-01031454.
PR   08-JUL-1999;   99DE-01031478.
PR   08-JUL-1999;   99DE-01031563.
PR   09-JUL-1999;   99DE-01032122.
PR   09-JUL-1999;   99DE-01032124.
PR   09-JUL-1999;   99DE-01032125.
PR   09-JUL-1999;   99DE-01032128.
PR   09-JUL-1999;   99DE-01032180.
PR   09-JUL-1999;   99DE-01032182.
PR   09-JUL-1999;   99DE-01032190.
PR   09-JUL-1999;   99DE-01032191.
PR   09-JUL-1999;   99DE-01032209.
PR   09-JUL-1999;   99DE-01032212.
PR   09-JUL-1999;   99DE-01032227.
PR   09-JUL-1999;   99DE-01032228.
PR   09-JUL-1999;   99DE-01032229.
PR   09-JUL-1999;   99DE-01032230.
PR   14-JUL-1999;   99DE-01032927.
PR   14-JUL-1999;   99DE-01033005.
PR   14-JUL-1999;   99DE-01033006.
PR   27-AUG-1999;   99DE-01040764.
PR   27-AUG-1999;   99DE-01040765.
PR   27-AUG-1999;   99DE-01040766.
PR   27-AUG-1999;   99DE-01040830.
PR   27-AUG-1999;   99DE-01040831.
PR   27-AUG-1999;   99DE-01040832.
PR   27-AUG-1999;   99DE-01040833.
PR   31-AUG-1999;   99DE-01041378.
PR   31-AUG-1999;   99DE-01041379.
PR   31-AUG-1999;   99DE-01041395.
PR   03-SEP-1999;   99DE-01042077.
PR   03-SEP-1999;   99DE-01042078.
PR   03-SEP-1999;   99DE-01042079.
PR   03-SEP-1999;   99DE-01042088.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-071486/08.
DR   N-PSDB; AAF67885.
XX
PT   Corynebacteium glutamicum nucleic acids encoding membrane construction
PT   and membrane transport proteins or their portions, useful for typing or
PT   identifying C. glutamicum or related bacteria, and as markers for
PT   transformation.
XX
PS   Claim 20; Page 571-572; 1119pp; English.
XX
CC   AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC   construction and membrane transport (MCT) proteins given in AAB76510 to
```

CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention
XX
SQ Sequence 425 AA;

Query Match 72.7%; Score 40; DB 4; Length 425;
Best Local Similarity 46.2%; Pred. No. 1.3e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 XKXVWANTLKAAX 13
:|:|| ||::| :
Db 296 DKSVWQNTIEACA 308

RESULT 8
AAB76651
ID AAB76651 standard; protein; 425 AA.
XX
AC AAB76651;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:284.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB000926.
XX
PR 25-JUN-1999; 99US-0141031P.
PR 08-JUL-1999; 99DE-01031454.
PR 08-JUL-1999; 99DE-01031478.
PR 08-JUL-1999; 99DE-01031563.
PR 09-JUL-1999; 99DE-01032122.
PR 09-JUL-1999; 99DE-01032124.
PR 09-JUL-1999; 99DE-01032125.
PR 09-JUL-1999; 99DE-01032128.
PR 09-JUL-1999; 99DE-01032180.
PR 09-JUL-1999; 99DE-01032182.
PR 09-JUL-1999; 99DE-01032190.
PR 09-JUL-1999; 99DE-01032191.
PR 09-JUL-1999; 99DE-01032209.
PR 09-JUL-1999; 99DE-01032212.
PR 09-JUL-1999; 99DE-01032227.
PR 09-JUL-1999; 99DE-01032228.
PR 09-JUL-1999; 99DE-01032229.
PR 09-JUL-1999; 99DE-01032230.
PR 14-JUL-1999; 99DE-01032927.
PR 14-JUL-1999; 99DE-01033005.
PR 14-JUL-1999; 99DE-01033006.
PR 27-AUG-1999; 99DE-01040764.
PR 27-AUG-1999; 99DE-01040765.
PR 27-AUG-1999; 99DE-01040766.
PR 27-AUG-1999; 99DE-01040830.
PR 27-AUG-1999; 99DE-01040831.
PR 27-AUG-1999; 99DE-01040832.
PR 27-AUG-1999; 99DE-01040833.
PR 31-AUG-1999; 99DE-01041378.
PR 31-AUG-1999; 99DE-01041379.
PR 31-AUG-1999; 99DE-01041395.
PR 03-SEP-1999; 99DE-01042077.
PR 03-SEP-1999; 99DE-01042078.
PR 03-SEP-1999; 99DE-01042079.
PR 03-SEP-1999; 99DE-01042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-071486/08.
DR N-PSDB; AAF67884.
XX
PT Corynebacteium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation.
XX
PS Claim 20; Page 568-569; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention
XX
SQ Sequence 425 AA;

Query Match 72.7%; Score 40; DB 4; Length 425;
Best Local Similarity 46.2%; Pred. No. 1.3e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 XKXVWANTLKAAX 13
:|:|| ||::| :
Db 296 DKSVWQNTIEACA 308

RESULT 9
AAB80185
ID AAB80185 standard; protein; 425 AA.
XX
AC AAB80185;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:1104.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB000923.
XX
PR 25-JUN-1999; 99US-0141031P.
PR 01-JUL-1999; 99DE-01030476.
PR 02-JUL-1999; 99US-0142101P.
PR 08-JUL-1999; 99DE-01031415.
PR 08-JUL-1999; 99DE-01031418.
PR 08-JUL-1999; 99DE-01031419.

```
R   08-JUL-1999;   99DE-01031420.
R   08-JUL-1999;   99DE-01031424.
R   08-JUL-1999;   99DE-01031428.
R   08-JUL-1999;   99DE-01031434.
R   08-JUL-1999;   99DE-01031435.
R   08-JUL-1999;   99DE-01031443.
R   08-JUL-1999;   99DE-01031453.
R   08-JUL-1999;   99DE-01031457.
R   08-JUL-1999;   99DE-01031465.
R   08-JUL-1999;   99DE-01031478.
R   08-JUL-1999;   99DE-01031510.
R   08-JUL-1999;   99DE-01031541.
R   08-JUL-1999;   99DE-01031573.
R   08-JUL-1999;   99DE-01031592.
R   08-JUL-1999;   99DE-01031632.
R   08-JUL-1999;   99DE-01031634.
R   08-JUL-1999;   99DE-01031636.
R   09-JUL-1999;   99DE-01032125.
R   09-JUL-1999;   99DE-01032126.
R   09-JUL-1999;   99DE-01032130.
R   09-JUL-1999;   99DE-01032186.
R   09-JUL-1999;   99DE-01032206.
R   09-JUL-1999;   99DE-01032227.
R   09-JUL-1999;   99DE-01032228.
R   09-JUL-1999;   99DE-01032229.
R   09-JUL-1999;   99DE-01032230.
R   14-JUL-1999;   99DE-01032922.
R   14-JUL-1999;   99DE-01032926.
R   14-JUL-1999;   99DE-01032928.
R   14-JUL-1999;   99DE-01033004.
R   14-JUL-1999;   99DE-01033005.
R   14-JUL-1999;   99DE-01033006.
R   12-AUG-1999;   99US-0148613P.
R   27-AUG-1999;   99DE-01040764.
R   27-AUG-1999;   99DE-01040765.
R   27-AUG-1999;   99DE-01040766.
R   27-AUG-1999;   99DE-01040832.
R   31-AUG-1999;   99DE-01041378.
R   31-AUG-1999;   99DE-01041379.
R   31-AUG-1999;   99DE-01041380.
R   31-AUG-1999;   99DE-01041394.
R   31-AUG-1999;   99DE-01041396.
R   03-SEP-1999;   99DE-01042076.
R   03-SEP-1999;   99DE-01042077.
R   03-SEP-1999;   99DE-01042079.
R   03-SEP-1999;   99DE-01042086.
R   03-SEP-1999;   99DE-01042087.
R   03-SEP-1999;   99DE-01042088.
R   03-SEP-1999;   99DE-01042095.
R   03-SEP-1999;   99DE-01042124.
R   03-SEP-1999;   99DE-01042129.
R   09-MAR-2000; 2000US-0187970P.
X
A   (BADI ) BASF AG.
X
I   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
X
R   WPI; 2001-137957/14.
R   N-PSDB; AAF72304.
X
T   Nucleic acids from Corynebacterium glutamicum encoding metabolic pathway
T   proteins, useful for producing fine chemicals in microorganisms,
T   including organic acids, nonproteinogenic amino acids, and purine and
T   pyrimidine bases.
X
S   Claim 20; Page 1662-1664; 1737pp; English.
X
C   AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
C   pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum MP
C   nucleic acids are useful for the production of fine chemicals in
C   microorganisms, including organic acids, nonproteinogenic amino acids,
C   purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated
C   and unsaturated fatty acids, diols, carbohydrates, aromatic compounds,
CC  vitamins, cofactors, polyketides and enzymes
XX
SQ  Sequence 425 AA;

  Query Match             72.7%;  Score 40;  DB 4;  Length 425;
  Best Local Similarity   46.2%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|:|| ||::| :
Db        296 DKSVWQNTIEACA 308


RESULT 10
AAR80504
ID  AAR80504 standard; protein; 537 AA.
XX
AC  AAR80504;
XX
DT  25-MAR-2003  (revised)
DT  04-DEC-1995  (first entry)
XX
DE  S. lividans protease Tap.
XX
KW  Protease; metalloendoproteinase; tripeptidyl aminopeptidase;
KW  protease-deficiency; protein secretion.
XX
OS  Streptomyces lividans.
XX
FH  Key             Location/Qualifiers
FT  Peptide         1. .36
FT                  /label= Sig_peptide
FT  Modified-site   1
FT                  /label= OTHER
FT                  /note= "fMet"
FT  Peptide         37. .39
FT                  /label= Autocatalytic_tripeptide
FT  Protein         40. .537
FT                  /label= Mat_protein
XX
PN  WO9517512-A2.
XX
PD  29-JUN-1995.
XX
PF  22-DEC-1994;   94WO-US014772.
XX
PR  23-DEC-1993;   93US-00173508.
XX
PA  (CANG-) CANGENE CORP.
XX
PI  Bartfeld D,  Butler MJ,  Hadary D,  Jenish DL,  Krieger TJ,  Malek LT;
PI  Walcyzk E,  Soostmeyer G;
XX
DR  WPI; 1995-240673/31.
DR  N-PSDB; AAQ99364.
XX
PT  Endogenous Streptomyces protease(s), opt. having impaired activity -
PT  useful in prodn. of exogenous proteins with reduced proteolytic
PT  degradation.
XX
PS  Claim 4; Fig 5; 142pp; English.
XX
CC  A genomic library of S. lividans 66 was prepd. in pSS12, and recombinants
CC  were used to transform S. lividans 66 protoplasts. Colonies selected for
CC  color formation on GPL-beta-naphthylamide contained a gene encoding a
CC  novel tripeptidyl aminopeptidase, Tap. Impaired expression of Tap by
CC  Streptomyces hosts improves the quality, quantity and stability of
CC  exogenous gene products. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ  Sequence 537 AA;

  Query Match             72.7%;  Score 40;  DB 2;  Length 537;
  Best Local Similarity   63.6%;  Pred. No. 1.7e+02;
```

```
 Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

y          1 XKXVWANTLKA 11
             : :||||| ||
o        142 KSAVWANTAKA 152


ESULT 11
AW87796
)   AAW87796 standard; protein; 537 AA.
(
:   AAW87796;
(
[   19-MAR-1999  (first entry)
(
E   Tripeptidyl aminopeptidase (tap) protein.
(
N   Tripeptidyl aminopeptidase; TAP; N-terminal cleavage; protein production;
N   GM-CSF; interleukin-3; IL-3; IL-6; EPO; tumour necrosis factor; TNF; SCF;
N   IL-7; IL-2.
(
S   Streptomyces lividans.
(
F   Key             Location/Qualifiers
[   Peptide         1. .39
[                   /note= "signal peptide"
[   Misc-difference 1
:                   /note= "Met encoded by TTG"
:   Protein         40. .537
:                   /note= "mature protein"
(
R   US5856166-A.
(
)   05-JAN-1999.
(
?   24-JUN-1994;   94US-00265310.
(
R   23-DEC-1993;   93US-00173508.
(
A   (CANG-) CANGENE CORP.

    Bartfeld D,  Malek LT,  Jenish DL,  Walczyk E,  Hadary D,  Garven S;
    Soostmeyer G,  Butler MJ,  Krygsman P,  Krieger TJ;
(
R   WPI; 1999-105117/09.
R   N-PSDB; AAV84065.

    Streptomyces tri:peptidyl aminopeptidase - useful for removing N-terminal
    pro-peptide from secreted proteins.

    Claim 2; Fig 12A-B; 83pp; English.

    The present sequence represents a tripeptidyl aminopeptidase (TAP) of
    Streptomyces. The aminopeptidase is endogenous to Streptomyces and
    cleaves an N-terminal sequence of X-Pro-Y, where X is an aliphatic or
    hydroxy amino acid and Y is an aliphatic, hydroxy or sulphur-containing
    amino acid. The TAP of Streptomyces are useful in the production of
    proteins, such as GM-CSF, interleukin-3 (IL-3), IL-6, EPO, tumour
    necrosis factor (TNF) SCF, IL-7 and IL-2

)   Sequence 537 AA;

 Query Match             72.7%;  Score 40;  DB 2;  Length 537;
 Best Local Similarity   63.6%;  Pred. No. 1.7e+02;
 Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

'          1 XKXVWANTLKA 11
             : :||||| ||
)        142 KSAVWANTAKA 152


ESULT 12
AB19228
```

```
ID   AAB19228 standard; protein; 537 AA.
XX
AC   AAB19228;
XX
DT   19-FEB-2001  (first entry)
XX
DE   A tripeptidyl aminopeptidase (TAP) polypeptide.
XX
KW   Tripeptidyl aminopeptidase; TAP; protein production; GM-CSF;
KW   chloromethylketone aminopeptidase inhibitor; stem cell factor;
KW   granulocyte macrophage-colony stimulating factor; interleukin-3; IL-3;
KW   IL-6; erythropoietin; EPO; SCF; IL-7; IL-2.
XX
OS   Streptomyces lividans.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .39
FT                   /note= "signal peptide"
FT   Modified-site   1
FT                   /note= "fMet"
FT   Protein         40. .537
FT                   /note= "mature protein"
XX
PN   US6127144-A.
XX
PD   03-OCT-2000.
XX
PF   16-OCT-1997;   97US-00951742.
XX
PR   23-DEC-1993;   93US-00173508.
PR   24-JUN-1994;   94US-00265310.
XX
PA   (CANG-) CANGENE CORP.
XX
PI   Bartfeld D,  Butler MJ,  Hadary D,  Krieger TJ,  Malek LT;
PI   Soostmeyer G,  Krygsman P,  Garven S,  Walczyk E,  Jenish DL;
XX
DR   WPI; 2000-655618/63.
DR   N-PSDB; AAC61403.
XX
PT   Producing a heterologous protein such as interleukins or growth factors
PT   involves incubating a transformed Streptomyces host cell in the presence
PT   of peptide-substituted chloromethylketone aminopeptidase inhibitor.
XX
PS   Disclosure; Fig 12A-C; 87pp; English.
XX
CC   The present sequence represents a tripeptidyl aminopeptidase (TAP)
CC   polypeptide. The polypeptide was used to identify inhibitors, which were
CC   then used in the method of the invention. The specification describes a
CC   method for producing a heterologous protein. The method involves
CC   incubating a Streptomyces host cell transformed nucleic acid sequence
CC   encoding the heterologous protein, in the presence of a peptide-
CC   substituted chloromethylketone aminopeptidase inhibitor. The inhibitor
CC   has the structure X-proline-Y-chlromethylketone, where X is an aliphatic
CC   or hydroxy amino acid and Y is an aliphatic hydroxy or sulphur-containing
CC   amino acid. Alternately, X and Y are non-polar amino acids. Use of the
CC   inhibitor inhibits degradation of the heterologous protein by
CC   aminopeptidases. The method is useful for producing granulocyte
CC   macrophage-colony stimulating factor (GM-CSF), interleukin-3 (IL-3), IL-
CC   6, erythropoietin (EPO), stem cell factor (SCF), IL-7, and IL-2 which are
CC   secreted from the host cell
XX
SQ   Sequence 537 AA;

  Query Match             72.7%;  Score 40;  DB 3;  Length 537;
  Best Local Similarity   63.6%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKA 11
              : :||||| ||
Db        142 KSAVWANTAKA 152
```

```
ESULT 13
BB07853
D   ABB07853 standard; protein; 541 AA.
X
C   ABB07853;
X
T   03-JUL-2002  (first entry)
X
E   C. glutamicum ABC transporter protein, Atr43.
X
W   ABC transporter protein; atr43; coryneform bacterium; fermentation;
W   L-amino acid; medicine; food; pharmaceutical; animal nutrition.
K
S   Corynebacterium glutamicum.
K
H   Key             Location/Qualifiers
T   Misc-difference 1
T                   /note= "encoded by TTG"
X
N   WO200222814-A2.
X
D   21-MAR-2002.
K
F   26-JUL-2001; 2001WO-EP008650.
K
R   15-SEP-2000; 2000DE-01045580.
R   11-MAY-2001; 2001DE-01023070.
K
A   (DEGS ) DEGUSSA AG.
X
I   Farwick M,  Huthmacher K,  Pfefferle W;
X
R   WPI; 2002-339870/37.
R   N-PSDB; ABL40819.
X
T   New atr43 gene of coryneform bacteria, useful when suppressed for
T   increasing fermentative production of L-amino acids, encodes an ABC
T   transporter protein.
X
S   Claim 7; Page 36-39; 41pp; English.
X
C   The invention relates to the atr43 gene from coryneform bacteria. The
C   encoded polypeptides have the activity of the ABC transporter protein
C   Atr43. Coryneforms having reduced expression of the atr43 gene are useful
C   for fermentative production of L-amino acids, specifically L-lysine,
C   useful in human medicine, the food and pharmaceutical industries and
C   particularly in animal nutrition. The atr43 polynucleotides are also
C   useful, as hybridisation probes or amplification primers, for identifying
C   nucleic acid that encodes the ABC transporter Atr43 and sequences closely
C   related to the atr43 gene, particularly where used as (micro)arrays or
C   DNA chips. The present sequence represents the Atr43 polypeptide
X
Q   Sequence 541 AA;

 Query Match             72.7%;  Score 40;  DB 5;  Length 541;
 Best Local Similarity   46.2%;  Pred. No. 1.7e+02;
 Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

y          1 XKXVWANTLKAAX 13
             :|:|| ||::| :
b        412 DKSVWQNTIEACA 424


ESULT 14
AG91463
D   AAG91463 standard; protein; 543 AA.
X
C   AAG91463;
X
T   26-SEP-2001  (first entry)
X
E   C glutamicum protein fragment SEQ ID NO: 5217.
X
KW  Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW  organic acid synthesis.
XX
OS  Corynebacterium glutamicum.
XX
PN  EP1108790-A2.
XX
PD  20-JUN-2001.
XX
PF  18-DEC-2000; 2000EP-00127688.
XX
PR  16-DEC-1999;   99JP-00377484.
PR  07-APR-2000; 2000JP-00159162.
PR  03-AUG-2000; 2000JP-00280988.
XX
PA  (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI  Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI  Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR  WPI; 2001-376931/40.
DR  N-PSDB; AAH66682.
XX
PT  Novel polynucleotides derived from Coryneform bacteria, for identifying
PT  mutation point of a gene, measuring expression of a gene, analyzing
PT  expression profile or pattern of a gene and identifying homologous gene.
XX
PS  Claim 17; SEQ ID NO 5217; 246pp + Sequence Listing; English.
XX
CC  The present invention provides a number of nucleotide and protein
CC  sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC  are useful for identifying the mutation point of a gene derived from a
CC  mutant of coryneform bacterium, measuring expression amount and analysing
CC  the expression profile or expression pattern of a gene derived from
CC  Coryneform bacterium, and identifying a homologue of a gene derived from
CC  coryneform bacterium. Coryneform bacteria are useful for producing amino
CC  acids, nucleic acids, vitamins, saccharides and organic acids,
CC  particularly L-lysine. The present sequence is a protein described in the
CC  exemplification of the invention. Note: The sequence data for this patent
CC  did not form part of the printed specification, but was obtained in
CC  electronic format directly from the European Patent Office
XX
SQ  Sequence 543 AA;

  Query Match             72.7%;  Score 40;  DB 4;  Length 543;
  Best Local Similarity   46.2%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|:|| ||::| :
Db        414 DKSVWQNTIEACA 426


RESULT 15
AAG10400
ID  AAG10400 standard; protein; 361 AA.
XX
AC  AAG10400;
XX
DT  17-OCT-2000  (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 8706.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
```

F 25-FEB-2000; 2000EP-00301439.
K
R 25-FEB-1999; 99US-0121825P.
R 05-MAR-1999; 99US-0123180P.
R 09-MAR-1999; 99US-0123548P.
R 23-MAR-1999; 99US-0125788P.
R 25-MAR-1999; 99US-0126264P.
R 29-MAR-1999; 99US-0126785P.
R 01-APR-1999; 99US-0127462P.
R 06-APR-1999; 99US-0128234P.
R 08-APR-1999; 99US-0128714P.
R 16-APR-1999; 99US-0129845P.
R 19-APR-1999; 99US-0130077P.
R 21-APR-1999; 99US-0130449P.
R 23-APR-1999; 99US-0130510P.
R 23-APR-1999; 99US-0130891P.
R 28-APR-1999; 99US-0131449P.
R 30-APR-1999; 99US-0132048P.
R 30-APR-1999; 99US-0132407P.
R 04-MAY-1999; 99US-0132484P.
R 05-MAY-1999; 99US-0132485P.
R 06-MAY-1999; 99US-0132486P.
R 06-MAY-1999; 99US-0132487P.
R 07-MAY-1999; 99US-0132863P.
R 11-MAY-1999; 99US-0134256P.
R 14-MAY-1999; 99US-0134218P.
R 14-MAY-1999; 99US-0134219P.
R 14-MAY-1999; 99US-0134221P.
R 14-MAY-1999; 99US-0134370P.
R 18-MAY-1999; 99US-0134768P.
R 19-MAY-1999; 99US-0134941P.
R 20-MAY-1999; 99US-0135124P.
R 21-MAY-1999; 99US-0135353P.
R 24-MAY-1999; 99US-0135629P.
R 25-MAY-1999; 99US-0136021P.
R 27-MAY-1999; 99US-0136392P.
R 28-MAY-1999; 99US-0136782P.
R 01-JUN-1999; 99US-0137222P.
R 03-JUN-1999; 99US-0137528P.
R 04-JUN-1999; 99US-0137502P.
R 07-JUN-1999; 99US-0137724P.
R 08-JUN-1999; 99US-0138094P.
R 10-JUN-1999; 99US-0138540P.
R 10-JUN-1999; 99US-0138847P.
R 14-JUN-1999; 99US-0139119P.
R 16-JUN-1999; 99US-0139452P.
R 16-JUN-1999; 99US-0139453P.
R 17-JUN-1999; 99US-0139492P.
R 18-JUN-1999; 99US-0139454P.
R 18-JUN-1999; 99US-0139455P.
R 18-JUN-1999; 99US-0139456P.
R 18-JUN-1999; 99US-0139457P.
R 18-JUN-1999; 99US-0139458P.
R 18-JUN-1999; 99US-0139459P.
R 18-JUN-1999; 99US-0139460P.
R 18-JUN-1999; 99US-0139461P.
R 18-JUN-1999; 99US-0139462P.
R 18-JUN-1999; 99US-0139463P.
R 18-JUN-1999; 99US-0139750P.
R 18-JUN-1999; 99US-0139763P.
R 21-JUN-1999; 99US-0139817P.
R 22-JUN-1999; 99US-0139899P.
R 23-JUN-1999; 99US-0140353P.
R 23-JUN-1999; 99US-0140354P.
R 24-JUN-1999; 99US-0140695P.
R 28-JUN-1999; 99US-0140823P.
R 29-JUN-1999; 99US-0140991P.
R 30-JUN-1999; 99US-0141287P.
R 01-JUL-1999; 99US-0141842P.
R 01-JUL-1999; 99US-0142154P.
R 02-JUL-1999; 99US-0142055P.
R 06-JUL-1999; 99US-0142390P.
R 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

R 28-SEP-1999; 99US-0156458P.
R 29-SEP-1999; 99US-0156596P.
R 04-OCT-1999; 99US-0157117P.
R 05-OCT-1999; 99US-0157753P.
R 06-OCT-1999; 99US-0157865P.
R 07-OCT-1999; 99US-0158029P.
R 08-OCT-1999; 99US-0158232P.
R 12-OCT-1999; 99US-0158369P.
R 13-OCT-1999; 99US-0159293P.
R 13-OCT-1999; 99US-0159294P.
R 13-OCT-1999; 99US-0159295P.
R 14-OCT-1999; 99US-0159329P.
R 14-OCT-1999; 99US-0159330P.
R 14-OCT-1999; 99US-0159331P.
R 14-OCT-1999; 99US-0159637P.
R 14-OCT-1999; 99US-0159638P.
R 18-OCT-1999; 99US-0159584P.
R 21-OCT-1999; 99US-0160741P.
R 21-OCT-1999; 99US-0160767P.
R 21-OCT-1999; 99US-0160768P.
R 21-OCT-1999; 99US-0160770P.
R 21-OCT-1999; 99US-0160814P.
R 21-OCT-1999; 99US-0160815P.
R 22-OCT-1999; 99US-0160980P.
R 22-OCT-1999; 99US-0160981P.
R 22-OCT-1999; 99US-0160989P.
R 25-OCT-1999; 99US-0161404P.
R 25-OCT-1999; 99US-0161405P.
R 25-OCT-1999; 99US-0161406P.
R 26-OCT-1999; 99US-0161359P.
R 26-OCT-1999; 99US-0161360P.
R 26-OCT-1999; 99US-0161361P.
R 28-OCT-1999; 99US-0161920P.
R 28-OCT-1999; 99US-0161992P.
R 28-OCT-1999; 99US-0161993P.
R 29-OCT-1999; 99US-0162142P.

Query Match 70.9%; Score 39; DB 3; Length 361;
Best Local Similarity 46.2%; Pred. No. 1.6e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```
y          1 XKXVWANTLKAAX 13
               :|::|||:  ||:
b        152 AKRIWANSQSAAR 164
```

earch completed: May 17, 2004, 13:49:39
ob time : 55 secs

4 protein - protein search, using sw model

ın on: May 17, 2004, 13:42:09 ; Search time 21 Seconds
(without alignments)
59.547 Million cell updates/sec

.tle: US-09-458-299A-4226
:rfect score: 55
:quence: 1 XKXVWANTLKAAX 13

:oring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

:arched: 283366 seqs, 96191526 residues

›tal number of hits satisfying chosen parameters: 283366

.nimum DB seq length: 0
ıximum DB seq length: 2000000000

›st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

ıtabase : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

›sult No.	Score	% Query Match	Length	DB	ID	Description
1	40	72.7	1695	2	JE0084	voltage-gated sodi
2	39	70.9	111	2	B75605	hypothetical prote
3	39	70.9	215	2	AH3215	hypothetical prote
4	38	69.1	239	2	AC2745	glycerophosphoryl
5	38	69.1	246	2	B97526	hypothetical prote
6	38	69.1	485	2	E86506	hypothetical prote
7	38	69.1	485	2	E72115	hypothetical prote
8	38	69.1	485	2	A81555	hypothetical prote
9	38	69.1	538	1	T40151	histidine-tRNA lig
10	38	69.1	1628	2	E90538	hypothetical prote
11	37	67.3	301	2	S51132	ADP,ATP carrier pr
12	37	67.3	357	2	AF2796	lipoprotein [impor
13	37	67.3	368	2	T03580	probable transcrip
14	37	67.3	371	2	F97575	hypothetical prote
15	37	67.3	384	2	D75201	hypothetical prote
16	37	67.3	398	2	S21883	bZIP transcription
17	37	67.3	423	2	H86195	hypothetical prote
18	37	67.3	1355	2	T32092	hypothetical prote
19	37	67.3	1739	2	A48298	sodium channel hom
20	36	65.5	111	2	B70035	chaperonin homolog
21	36	65.5	162	2	AF1435	PTS system, fructo
22	36	65.5	162	2	AG1077	PTS system, fructo
23	36	65.5	216	2	A72291	hypothetical prote
24	36	65.5	303	2	S60550	envelope polyprote
25	36	65.5	303	2	S60549	envelope polyprote
26	36	65.5	337	2	E96543	probable RAV-like
27	36	65.5	339	2	A41677	ADP,ATP carrier pr
28	36	65.5	443	2	C41621	env polyprotein P
29	36	65.5	508	2	T20757	hypothetical prote
30	36	65.5	662	2	T01857	hypothetical prote
31	36	65.5	721	2	A43275	85K protein 1(1)zw
32	36	65.5	765	1	S57636	5-methyltetrahydro
33	36	65.5	765	2	T12575	5-methyltetrahydro
34	36	65.5	1038	2	D95252	fucolectin-related
35	36	65.5	1038	2	B98117	hypothetical prote
36	36	65.5	1358	2	A29360	SIR4 protein - yea
37	35	63.6	125	2	B72536	hypothetical prote
38	35	63.6	134	1	G64061	virulence-associat
39	35	63.6	139	2	S76176	hypothetical prote
40	35	63.6	149	2	A70325	hypothetical prote
41	35	63.6	172	2	S28683	hypothetical prote
42	35	63.6	214	2	D86533	frame-shift with C
43	35	63.6	226	1	QQAG1T	hypothetical prote
44	35	63.6	246	2	AE3354	glycerophosphodies
45	35	63.6	270	2	G85078	probable transposo

ALIGNMENTS

RESULT 1
JE0084
voltage-gated sodium channel alpha subunit - hydromedusa (Polyorchis penicillatus)
N;Alternate names: PpSCN 1
C;Species: Polyorchis penicillatus
C;Date: 11-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: JE0084
R;Spafford, J.D.; Spencer, A.N.; Gallin, W.J.
Biochem. Biophys. Res. Commun. 244, 772-780, 1998
A;Title: A putative voltage-gated sodium channel alpha subunit (PpSCN1) from the hydrozc
A;Reference number: JE0084; MUID:98205797; PMID:9535741
A;Accession: JE0084
A;Molecule type: mRNA
A;Residues: 1-1695 <SPA>
A;Cross-references: GB:AF047380; NID:g3005563; PIDN:AAC38974.1; PID:g3005564
C;Comment: This protein is the only pore-forming alpha subunit available to account for
C;Superfamily: sodium channel protein
C;Keywords: glycoprotein
F;201,273,299,684,1065,1082,1089,1428/Binding site: carbohydrate (Asn) (covalent) #statu

Query Match 72.7%; Score 40; DB 2; Length 1695;
Best Local Similarity 77.8%; Pred. No. 77;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          5 WANTLKAAX 13
              | ||||||:
Db       1192 WVNTLKAAS 1200
```

RESULT 2
B75605
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C;Accession: B75605
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75605
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-111 <WHI>
A;Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12331.1; PID:g646062
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0104
A;Map position: 2
C;Superfamily: Deinococcus radiodurans hypothetical protein DRA0104

```
 Query Match             70.9%;  Score 39;  DB 2;  Length 111;
 Best Local Similarity   63.6%;  Pred. No. 6.3;
 Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

y          3 XVWANTLKAAX 13
             :|||:| ||:
b         62 SVWANSLDAAI 72


ESULT 3
H3215
ypothetical protein Atu5455 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
;Species: Agrobacterium tumefaciens
;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
;Accession: AH3215
;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
rage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Karp, P.; Romero, P.; Zhang, S.
cience 294, 2317-2323, 2001
;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ter, E.W.
;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
;Reference number: AB2577; MUID:21608550; PMID:11743193
;Accession: AH3215
;Status: preliminary
;Molecule type: DNA
;Residues: 1-215 <KUR>
;Cross-references: GB:AE008687; PIDN:AAL46142.1; PID:g17743910; GSPDB:GN00188
;Experimental source: strain C58 (Dupont)
;Genetics:
;Gene: Atu5455
;Genome: plasmid

 Query Match             70.9%;  Score 39;  DB 2;  Length 215;
 Best Local Similarity   63.6%;  Pred. No. 13;
 Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

y          3 XVWANTLKAAX 13
             :||| :||||:
b        194 AVWAQSLKAAL 204


ESULT 4
C2745
lycerophosphoryl diester phosphodiesterase Atu1371 [imported] - Agrobacterium tumefacie
;Species: Agrobacterium tumefaciens
;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
;Accession: AC2745
;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
rage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Karp, P.; Romero, P.; Zhang, S.
cience 294, 2317-2323, 2001
;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ter, E.W.
;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
;Reference number: AB2577; MUID:21608550; PMID:11743193
;Accession: AC2745
;Status: preliminary
;Molecule type: DNA
;Residues: 1-239 <KUR>
;Cross-references: GB:AE008688; PIDN:AAL42377.1; PID:g17739785; GSPDB:GN00186
;Experimental source: strain C58 (Dupont)
;Genetics:
;Gene: Atu1371
;Map position: circular chromosome

 Query Match             69.1%;  Score 38;  DB 2;  Length 239;
 Best Local Similarity   54.5%;  Pred. No. 22;
 Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

y          1 XKXVWANTLKA 11
             :|:|| ||: |
b         15 NKAVWENTISA 25


RESULT 5
B97526
hypothetical protein AGR_C_2533 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: B97526
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B97526
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87163.1; PID:g15156435; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_2533
A;Map position: circular chromosome

  Query Match             69.1%;  Score 38;  DB 2;  Length 246;
  Best Local Similarity   54.5%;  Pred. No. 23;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 XKXVWANTLKA 11
             :|:|| ||: |
Db        22 NKAVWENTISA 32


RESULT 6
E86506
hypothetical protein CPj0124 [imported] - Chlamydophila pneumoniae (strain J138)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: E86506
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349; PMID:10871362
A;Accession: E86506
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-485 <STO>
A;Cross-references: GB:BA000008; NID:g8978498; PIDN:BAA98335.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: CPj0124

  Query Match             69.1%;  Score 38;  DB 2;  Length 485;
  Best Local Similarity   63.6%;  Pred. No. 47;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         3 XVWANTLKAAX 13
             :|||| | ||:
Db       412 SVWANQLSAAE 422


RESULT 7
E72115
hypothetical protein - Chlamydophila pneumoniae (strain CWL029)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C;Accession: E72115
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606; PMID:10192388
A;Accession: E72115
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-485 <ARN>
```

```
Cross-references: GB:AE001599; GB:AE001363; NID:g4376387; PIDN:AAD18277.1; PID:g437638
Experimental source: strain CWL029
Genetics:
Gene: CPn0124

Query Match             69.1%;  Score 38;  DB 2;  Length 485;
Best Local Similarity   63.6%;  Pred. No. 47;
Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

          3 XVWANTLKAAX 13
            :|||| | ||:
        412 SVWANQLSAAE 422


SULT 8
1555
pothetical protein CP0649 [imported] - Chlamydophila pneumoniae (strain AR39)
Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
Accession: A81555
Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
cleic Acids Res. 28, 1397-1406, 2000
Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
Reference number: A81500; MUID:20150255; PMID:10684935
Accession: A81555
Status: preliminary
Molecule type: DNA
Residues: 1-485 <REA>
Cross-references: GB:AE002222; GB:AE002161; NID:g7189553; PIDN:AAF38464.1; PID:g718956
Experimental source: strain AR39, HL cells
Genetics:
Gene: CP0649

Query Match             69.1%;  Score 38;  DB 2;  Length 485;
Best Local Similarity   63.6%;  Pred. No. 47;
Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

          3 XVWANTLKAAX 13
            :|||| | ||:
        412 SVWANQLSAAE 422


SULT 9
0151
stidine-tRNA ligase precursor, mitochondrial - fission yeast (Schizosaccharomyces pomb
Species: Schizosaccharomyces pombe
Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-Jul-2002
Accession: T40151
Wood, V.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.
bmitted to the EMBL Data Library, March 1998
Reference number: Z21842
Accession: T40151
Status: preliminary; translated from GB/EMBL/DDBJ
Molecule type: DNA
Residues: 1-538 <WOO>
Cross-references: EMBL:AL022103; PIDN:CAA17892.1; GSPDB:GN00067; SPDB:SPBC2G2.12
Experimental source: strain 972h-; cosmid c2G2
Genetics:
Gene: SPDB:SPBC2G2.12
Map position: 2
Genome: nuclear
Superfamily: human histidine-tRNA ligase; amino acid-tRNA ligase repeat homology; hist
Keywords: mitochondrion
15-65/Domain: amino acid-tRNA ligase repeat homology <ATL>
66-526/Domain: histidine-tRNA ligase homology <HTL>

Query Match             69.1%;  Score 38;  DB 1;  Length 538;
Best Local Similarity   54.5%;  Pred. No. 52;
Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

          1 XKXVWANTLKA 11
            :|::||| :||
Db        457 CKDLWANGIKA 467


RESULT 10
E90538
hypothetical protein MYPU_2130 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: E90538
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: E90538
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1628 <KUR>
A;Cross-references: GB:AL445566; PID:g14089626; PIDN:CAC13386.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_2130
A;Genetic code: SGC3

  Query Match             69.1%;  Score 38;  DB 2;  Length 1628;
  Best Local Similarity   46.2%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|: | |||:| :
Db        916 VKNTWLNTLEATG 928


RESULT 11
S51132
ADP,ATP carrier protein - malaria parasite (Plasmodium falciparum)
N;Alternate names: ADP/ATP transporter
C;Species: Plasmodium falciparum
C;Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jun-2000
C;Accession: S68993; S51132
R;Hatin, I.; Jaureguiberry, G.
Eur. J. Biochem. 228, 86-91, 1995
A;Title: Molecular characterisation of the ADP/ATP-transporter cDNA from the human malar
A;Reference number: S68993; MUID:95188918; PMID:7883016
A;Accession: S68993
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-301 <HAT>
A;Cross-references: EMBL:X83551; NID:g623334; PIDN:CAA58541.1; PID:g623335
C;Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology
C;Keywords: duplication; transmembrane protein
F;6-102/Domain: ADP,ATP carrier protein repeat homology <ACP1>
F;112-203/Domain: ADP,ATP carrier protein repeat homology <ACP2>
F;209-301/Domain: ADP,ATP carrier protein repeat homology <ACP3>

  Query Match             67.3%;  Score 37;  DB 2;  Length 301;
  Best Local Similarity   38.5%;  Pred. No. 43;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|: ||| :: |:
Db        274 FKGAWANVIRGAG 286


RESULT 12
AF2796
lipoprotein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2796
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
```

;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ter, E.W.
;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
;Reference number: AB2577; MUID:21608550; PMID:11743193
;Accession: AF2796
;Status: preliminary
;Molecule type: DNA
;Residues: 1-357 <KUR>
;Cross-references: GB:AE008688; PIDN:AAL42788.1; PID:g17740232; GSPDB:GN00186
;Experimental source: strain C58 (Dupont)
;Genetics:
;Gene: Atu1789
;Map position: circular chromosome

Query Match 67.3%; Score 37; DB 2; Length 357;
Best Local Similarity 75.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

y 1 XKXVWANT 8
:|:|||||
b 184 LKVVWANT 191

ESULT 13
03580
robable transcription activator RF2a - rice
;Species: Oryza sativa (rice)
;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
;Accession: T03580
;Yin, Y.; Zhu, Q.; Dai, S.; Lamb, C.; Beachy, R.N.
MBO J. 16, 5247-5259, 1997
;Title: RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacil
;Reference number: Z14956; MUID:97459912; PMID:9311985
;Accession: T03580
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-368 <YIN>
;Cross-references: EMBL:AF005492; NID:g2253277; PIDN:AAC49832.1; PID:g2253278
;Experimental source: strain TP301
;Genetics:
;Note: rf2a

Query Match 67.3%; Score 37; DB 2; Length 368;
Best Local Similarity 46.2%; Pred. No. 53;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

y 1 XKXVWANTLKAAX 13
:|::||| ||:
b 173 AKRIWANRQSAAR 185

ESULT 14
97575
ypothetical protein AGR_C_3292 [imported] - Agrobacterium tumefaciens (strain C58, Cere
;Species: Agrobacterium tumefaciens
;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
;Accession: F97575
;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
cience 294, 2323-2328, 2001
;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
;Reference number: A97359; MUID:21608551; PMID:11743194
;Accession: F97575
;Status: preliminary
;Molecule type: DNA
;Residues: 1-371 <KUR>
;Cross-references: GB:AE007869; PIDN:AAK87559.1; PID:g15156897; GSPDB:GN00169
;Genetics:
;Gene: AGR_C_3292
;Map position: circular chromosome

Query Match 67.3%; Score 37; DB 2; Length 371;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XKXVWANT 8
:|:|||||
Db 198 LKVVWANT 205

RESULT 15
D75201
hypothetical protein PAB0085 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: D75201
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A;Reference number: A75001
A;Accession: D75201
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-384 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49059.1; PID:e151495
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0085

Query Match 67.3%; Score 37; DB 2; Length 384;
Best Local Similarity 63.6%; Pred. No. 55;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 XVWANTLKAAX 13
::|| |||||:
Db 167 ILWAVTLKAAW 177

Search completed: May 17, 2004, 13:51:28
Job time : 22 secs

GenCore version 5.1.6

protein - protein search, using sw model

n on: May 17, 2004, 13:33:59 ; Search time 11 Seconds
(without alignments)
61.537 Million cell updates/sec

tle: US-09-458-299A-4226
rfect score: 55
quence: 1 XKXVWANTLKAAX 13

oring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

tal number of hits satisfying chosen parameters: 141681

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID		Description
1	37	67.3	398	1	PF21_ARATH	Q04088	arabidopsis
2	37	67.3	509	1	SYK_ACICA	Q43990	acinetobact
3	36	65.5	111	1	YVDS_BACSU	O32262	bacillus su
4	36	65.5	335	1	OTC_STRCO	Q93jf1	streptomyce
5	36	65.5	339	1	ADT_CHLKE	P31692	chlorella k
6	36	65.5	357	1	IDI2_CHLTE	Q8kfr5	chlorobium
7	36	65.5	721	1	ZW10_DROME	Q9w4x9	drosophila
8	36	65.5	765	1	METE_CATRO	Q42699	catharanthu
9	36	65.5	765	1	METE_MESCR	P93263	mesembryant
10	36	65.5	1358	1	SIR4_YEAST	P11978	saccharomyc
11	35	63.6	134	1	Y322_HAEIN	Q57122	haemophilus
12	35	63.6	149	1	Y272_AQUAE	O66629	aquifex aeo
13	35	63.6	172	1	YP1_AGRT4	P04028	agrobacteri
14	35	63.6	404	1	CGE2_HUMAN	O96020	homo sapien
15	35	63.6	655	1	AMYA_PYRAB	Q9v298	pyrococcus
16	35	63.6	697	1	Y351_BUCAP	Q8k9i4	buchnera ap
17	35	63.6	790	1	YDDB_ECOLI	P31827	escherichia
18	35	63.6	1053	1	HMDH_SCHPO	Q10283	schizosacch
19	35	63.6	1061	1	CYPD_BACSU	O08394	bacillus su
20	34	61.8	70	1	RK28_CYAPA	P48129	cyanophora
21	34	61.8	99	1	Y11K_STRFR	P26800	streptomyce
22	34	61.8	176	1	FRH1_XENLA	P17663	xenopus lae
23	34	61.8	176	1	FRH2_XENLA	P49948	xenopus lae
24	34	61.8	194	1	H5B_XENLA	P22845	xenopus lae
25	34	61.8	211	1	HPRT_LEIDO	P43152	leishmania
26	34	61.8	317	1	RPC6_YEAST	P32910	saccharomyc
27	34	61.8	366	1	T2AA_DROME	P52654	drosophila
28	34	61.8	423	1	YDIN_ECOLI	P76198	escherichia
29	34	61.8	430	1	RT05_HUMAN	P82675	homo sapien
30	34	61.8	591	1	VATA_CHLMU	Q9pk85	chlamydia m
31	34	61.8	591	1	VATA_CHLTR	O84310	chlamydia t
32	34	61.8	633	1	AMYA_PYRHO	O57932	pyrococcus
33	34	61.8	653	1	MALQ_PYRKO	O32450	pyrococcus
34	34	61.8	659	1	MALQ_THELI	O32462	thermococcu
35	34	61.8	831	1	IF2_RICPR	Q9zcz8	rickettsia
36	34	61.8	1151	1	ATC7_YEAST	P40527	saccharomyc
37	34	61.8	1518	1	POLG_PPVEA	Q01681	p genome po
38	34	61.8	3066	1	POLG_SBMVG	Q90069	s genome po
39	34	61.8	3066	1	POLG_SBMVN	P21231	s genome po
40	34	61.8	3125	1	POLG_PPVNA	P17766	p genome po
41	34	61.8	3140	1	POLG_PPVRA	P17767	p genome po
42	34	61.8	3140	1	POLG_PPVSK	Q84934	p genome po
43	34	61.8	3141	1	POLG_PPVD	P13529	p genome po
44	34	61.8	3206	1	POLG_PSBMV	P29152	p genome po
45	33	60.0	176	1	FRI1_RANCA	P07229	rana catesb

ALIGNMENTS

RESULT 1
PF21_ARATH
ID PF21_ARATH STANDARD; PRT; 398 AA.
AC Q04088;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Possible transcription factor PosF21.
GN POSF21 OR AT2G31370 OR T28P16.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Zurich; TISSUE=Leaf;
RX MEDLINE=93251100; PubMed=1844885;
RA Aeschbacher R.A., Schrott M., Potrykus I., Saul M.W.;
RT "Isolation and molecular characterization of PosF21, an Arabidopsis
RT thaliana gene which shows characteristics of a b-Zip class
RT transcription factor.";
RL Plant J. 1:303-316(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";

```
:L   Science 302:842-846(2003).
:C   -!- FUNCTION: Putative transcription factor with an activatory role.
:C   -!- SUBCELLULAR LOCATION: Nuclear.
:C   -!- ALTERNATIVE PRODUCTS:
:C       Event=Alternative splicing; Named isoforms=1;
:C         Comment=A number of isoforms are produced. According to EST
:C         sequences;
:C       Name=1;
:C         IsoId=Q04088-1; Sequence=Displayed;
:C   -!- DEVELOPMENTAL STAGE: Expressed constitutively at a low level in
:C       young seedlings and in roots, stems and leaves of mature plants.
:C   -!- SIMILARITY: Belongs to the bZIP family.
C    ---------------------------------------------------------------------
C    This SWISS-PROT entry is copyright. It is produced through a collaboration
C    between the Swiss Institute of Bioinformatics and the EMBL outstation -
C    the European Bioinformatics Institute. There are no restrictions on its
C    use by non-profit institutions as long as its content is in no way
C    modified and this statement is not removed. Usage by and for commercial
C    entities requires a license agreement (See http://www.isb-sib.ch/announce/
C    or send an email to license@isb-sib.ch).
C    ---------------------------------------------------------------------
R    EMBL; X61031; CAA43366.1; -.
R    EMBL; AC007169; AAD26486.1; -.
R    EMBL; AY057534; AAL09774.1; -.
R    EMBL; AY113058; AAM47366.1; -.
R    PIR; S21883; S21883.
R    InterPro; IPR004827; TF_bZIP.
R    Pfam; PF00170; bZIP; 1.
R    SMART; SM00338; BRLZ; 1.
R    PROSITE; PS50217; BZIP; 1.
R    PROSITE; PS00036; BZIP_BASIC; FALSE_NEG.
W    Transcription regulation; Activator; Nuclear protein; DNA-binding;
W    Alternative splicing.
T    DNA_BIND     203     222       BASIC MOTIF.
T    DOMAIN       229     264       LEUCINE-ZIPPER.
T    DOMAIN       340     372       POLY-GLN.
Q    SEQUENCE   398 AA;  44689 MW;  2DAA9EC9B9C14D11 CRC64;

 Query Match             67.3%;  Score 37;  DB 1;  Length 398;
 Best Local Similarity   46.2%;  Pred. No. 22;
 Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|::|||   ||:
Db        205 AKRIWANRQSAAR 217


ESULT 2
YK_ACICA
D    SYK_ACICA      STANDARD;      PRT;   509 AA.
C    Q43990;
T    01-NOV-1997 (Rel. 35, Created)
T    01-NOV-1997 (Rel. 35, Last sequence update)
T    28-FEB-2003 (Rel. 41, Last annotation update)
E    Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).
N    LYSS.
S    Acinetobacter calcoaceticus.
C    Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
C    Moraxellaceae; Acinetobacter.
X    NCBI_TaxID=471;
N    [1]
P    SEQUENCE FROM N.A.
C    STRAIN=BD413 / ADP1;
X    MEDLINE=97228433; PubMed=9074511;
A    Geissdoerfer W., Ratajczak A., Hillen W.;
T    "Nucleotide sequence of a putative periplasmic Mn superoxide dismutase
T    from Acinetobacter calcoaceticus ADP1.";
L    Gene 186:305-308(1997).
C    -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
C        + L-lysyl-tRNA(Lys).
C    -!- COFACTOR: Binds 3 magnesium ions per subunit (By similarity).
C    -!- SUBUNIT: Homodimer (By similarity).
C    -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46863; CAA86924.1; -.
DR   HSSP; P14825; 1E1O.
DR   HAMAP; MF_00252; -; 1.
DR   InterPro; IPR008994; Nucleic_acid_OB.
DR   InterPro; IPR004364; tRNA-synt_2.
DR   InterPro; IPR002313; tRNA-synt_lys_2.
DR   InterPro; IPR004365; tRNA_anti.
DR   InterPro; IPR006195; tRNA_ligase_II.
DR   Pfam; PF00152; tRNA-synt_2; 1.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   PRINTS; PR00982; TRNASYNTHLYS.
DR   TIGRFAMs; TIGR00499; lysS_bact; 1.
DR   PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Metal-binding; Magnesium.
FT   METAL       418     418       MAGNESIUM 1 (BY SIMILARITY).
FT   METAL       425     425       MAGNESIUM 1 AND 2 (BY SIMILARITY).
SQ   SEQUENCE   509 AA;  58079 MW;  95ED1AA43DC3D2F6 CRC64;

  Query Match             67.3%;  Score 37;  DB 1;  Length 509;
  Best Local Similarity   60.0%;  Pred. No. 29;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLK 10
              :|:|| || |
Db         38 GKSVWPNTFK 47


RESULT 3
YVDS_BACSU
ID   YVDS_BACSU     STANDARD;      PRT;   111 AA.
AC   O32262; O07000;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein yvdS.
GN   YVDS OR BSU34490.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Denizot F.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
```

```
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
Nature 390:249-256(1997).
-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
-!- SIMILARITY: Belongs to the small multidrug resistance (SMR)
    protein family.
----------------------------------------------------------------------
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
----------------------------------------------------------------------
EMBL; Z94043; CAB08048.1; ALT_INIT.
EMBL; Z99121; CAB15454.1; -.
PIR; B70035; B70035.
SubtiList; BG12427; yvdS.
InterPro; IPR000390; Smr.
Pfam; PF00893; Multi_Drug_Res; 1.
Hypothetical protein; Transmembrane; Transport; Complete proteome.
TRANSMEM      3     23       POTENTIAL.
TRANSMEM     24     44       POTENTIAL.
TRANSMEM     54     74       POTENTIAL.
TRANSMEM     80    100       POTENTIAL.
SEQUENCE   111 AA;  12085 MW;  655E7165743DFB6E CRC64;

 Query Match             65.5%;  Score 36;  DB 1;  Length 111;
 Best Local Similarity   46.2%;  Pred. No. 8.9;
 Matches    6;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

          1 XKXVWANTLKAAX 13
            :::|||::||  |:
         12 LEVVWASSLKHAD 24


SULT 4
C_STRCO
   OTC_STRCO      STANDARD;      PRT;   335 AA.
   Q93JF1;
   10-OCT-2003 (Rel. 42, Created)
   10-OCT-2003 (Rel. 42, Last sequence update)
   10-OCT-2003 (Rel. 42, Last annotation update)
   Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCase).
   ARGF OR ARCB OR SCO5976 OR STBAC16H6.11 OR SCBAC16H6.11.
   Streptomyces coelicolor.
   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
   Streptomycineae; Streptomycetaceae; Streptomyces.
   NCBI_TaxID=1902;
   [1]
   SEQUENCE FROM N.A.
   STRAIN=A3(2) / M145;
   MEDLINE=21996410; PubMed=12000953;
   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
   Hopwood D.A.;
   "Complete genome sequence of the model actinomycete Streptomyces
   coelicolor A3(2).";
   Nature 417:141-147(2002).
   -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine = phosphate
       + L-citrulline.
   -!- PATHWAY: Arginine biosynthesis; sixth step.
   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
   -!- SIMILARITY: Belongs to the ATCase/OTCase family.
   ----------------------------------------------------------------------
   This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
   ----------------------------------------------------------------------
   EMBL; AL596162; CAC44590.1; -.
   HAMAP; MF_01109; -; 1.
   InterPro; IPR006130; Asp/Orn_COtranf.
   InterPro; IPR002292; Orn_carbtransf.
   InterPro; IPR006131; OTCace_O.
   InterPro; IPR006132; OTCace_P.
   Pfam; PF00185; OTCace; 1.
   Pfam; PF02729; OTCace_N; 1.
   PRINTS; PR00100; AOTCASE.
   TIGRFAMs; TIGR00658; orni_carb_tr; 1.
   PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
   Arginine biosynthesis; Transferase; Complete proteome.
   SITE         35     35       IMPORTANT FOR STRUCTURAL INTEGRITY (BY
                                SIMILARITY).
   SITE         60     64       CARBAMOYLPHOSPHATE BINDING (BY
                                SIMILARITY).
   SITE        111    111       CARBAMOYLPHOSPHATE BINDING (BY
                                SIMILARITY).
   SITE        138    138       CARBAMOYLPHOSPHATE BINDING (BY
                                SIMILARITY).
   SITE        151    151       IMPORTANT FOR STRUCTURAL INTEGRITY (BY
                                SIMILARITY).
   SITE        276    279       ORNITHINE BINDING (BY SIMILARITY).
   SEQUENCE   335 AA;  36701 MW;  6FAFC3FBC876D337 CRC64;

    Query Match             65.5%;  Score 36;  DB 1;  Length 335;
    Best Local Similarity   54.5%;  Pred. No. 29;
    Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKA 11
              :|:|||  :||
Db        243 PKEVWAERIKA 253


RESULT 5
ADT_CHLKE
ID   ADT_CHLKE      STANDARD;      PRT;   339 AA.
AC   P31692;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   ADP,ATP carrier protein (ADP/ATP translocase) (Adenine nucleotide
DE   translocator) (ANT).
OS   Chlorella kessleri.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellales incertae sedis; Parachlorella.
OX   NCBI_TaxID=3074;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92084708; PubMed=1748677;
RA   Hilgarth C., Sauer N., Tanner W.;
RT   "Glucose increases the expression of the ATP/ADP translocator and the
```

```
T   glyceraldehyde-3-phosphate dehydrogenase genes in Chlorella.";
L   J. Biol. Chem. 266:24044-24047(1991).
C   -!- FUNCTION: Catalyzes the exchange of ADP and ATP across the
C       mitochondrial inner membrane.
C   -!- SUBUNIT: Homodimer (By similarity).
C   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
C       inner membrane.
C   -!- SIMILARITY: Belongs to the mitochondrial carrier family.
C   -!- SIMILARITY: Contains 3 Solcar repeats.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; M76669; AAA33027.1; -.
R   PIR; A41677; A41677.
R   InterPro; IPR002067; Mit_carrier.
R   InterPro; IPR001993; Mitoch_carrier.
R   Pfam; PF00153; mito_carr; 3.
R   PRINTS; PR00926; MITOCARRIER.
R   PROSITE; PS50920; SOLCAR; 3.
W   Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
F   TRANSMEM     45     62       1 (POTENTIAL).
F   TRANSMEM    108    126       2 (POTENTIAL).
F   TRANSMEM    151    168       3 (POTENTIAL).
F   TRANSMEM    209    228       4 (POTENTIAL).
F   TRANSMEM    248    265       5 (POTENTIAL).
F   TRANSMEM    304    322       6 (POTENTIAL).
F   REPEAT       39    133       SOLCAR 1.
F   REPEAT      145    234       SOLCAR 2.
F   REPEAT      246    328       SOLCAR 3.
Q   SEQUENCE   339 AA;  36686 MW;  54779734A33B3942 CRC64;

 Query Match             65.5%;  Score 36;  DB 1;  Length 339;
 Best Local Similarity   38.5%;  Pred. No. 29;
 Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

y          1 XKXVWANTLKAAX 13
             :|: |:| |: |:
b        302 FKGAWSNVLRGAG 314


ESULT 6
DI2_CHLTE
D   IDI2_CHLTE     STANDARD;      PRT;   357 AA.
C   Q8KFR5;
T   10-OCT-2003 (Rel. 42, Created)
T   10-OCT-2003 (Rel. 42, Last sequence update)
T   10-OCT-2003 (Rel. 42, Last annotation update)
E   Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
E   (Isopentenyl pyrophosphate isomerase).
N   FNI OR CT0257.
S   Chlorobium tepidum.
C   Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
C   Chlorobium.
X   NCBI_TaxID=1097;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=TLS / ATCC 49652 / DSM 12025;
X   MEDLINE=22103685; PubMed=12093901;
A   Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
A   Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
A   Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
A   Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
A   Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
A   Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
A   Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
T   "The complete genome sequence of Chlorobium tepidum TLS, a
T   photosynthetic, anaerobic, green-sulfur bacterium.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC  -!- FUNCTION: Catalyzes the 1,3-allylic rearrangement of the
CC      homoallylic substrate isopentenyl (IPP) to its allylic isomer,
CC      dimethylallyl diphosphate (DMAPP) (By similarity).
CC  -!- CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
CC      diphosphate.
CC  -!- COFACTOR: FMN and NADPH (By similarity).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC  -!- SIMILARITY: Belongs to the IPP isomerase type 2 family.
CC  ---------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; AE012804; AAM71503.1; -.
DR  TIGR; CT0257; -.
DR  HAMAP; MF_00354; -; 1.
DR  InterPro; IPR003009; FMN_enzyme.
KW  Isomerase; Isoprene biosynthesis; Flavoprotein; FMN; NADP;
KW  Complete proteome.
SQ  SEQUENCE   357 AA;  38265 MW;  4D2AE29D335C785C CRC64;

  Query Match             65.5%;  Score 36;  DB 1;  Length 357;
  Best Local Similarity   54.5%;  Pred. No. 31;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 XVWANTLKAAX 13
              : ||| |:||:
Db        327 RTWANDLRAAM 337


RESULT 7
ZW10_DROME
ID   ZW10_DROME     STANDARD;      PRT;   721 AA.
AC   Q9W4X9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Centromere/kinetochore protein zw10 (Mitotic 15 protein).
GN   MIT(1)15 OR ZW10 OR EG:BACR7C10.3 OR CG9900.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Imaginal disks;
RX   MEDLINE=92363920; PubMed=1339459;
RA   Williams B.C., Karr T.L., Montgomery J.M., Goldberg M.L.;
RT   "The Drosophila l(1)zw10 gene product, required for accurate mitotic
RT   chromosome segregation, is redistributed at anaphase onset.";
RL   J. Cell Biol. 118:759-773(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
```

```
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=20196011; PubMed=10731137;
RA   Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA   Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA   Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA   Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA   Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA   Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
RA   Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,
RA   Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA   McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA   Glover D.M.;
RT   "From sequence to chromosome: the tip of the X chromosome of D.
RT   melanogaster.";
RL   Science 287:2220-2222(2000).
CC   -!- FUNCTION: Required for accurate chromosome segregation.
CC   -!- SUBCELLULAR LOCATION: EXCLUDED FROM THE NUCLEUS DURING INTERPHASE
CC       BUT MIGRATES INTO THE NUCLEAR ZONE DURING PROMETAPHASE. AT
CC       METAPHASE, FOUND IN A FILAMENTOUS STRUCTURE THAT MAY BE
CC       SPECIFICALLY ASSOCIATED WITH KINETOCHORE MICROTUBULES. AT
CC       ANAPHASE, FOUND AT OR NEAR KINETOCHORES OF SEPARATING CHROMOSOMES.
CC       AT THE BEGINNING OF TELOPHASE, BECOMES EXCLUDED AGAIN FROM THE
CC       NUCLEUS AND IS DISPERSED IN THE CYTOPLASM.
CC   -!- DEVELOPMENTAL STAGE: HIGHEST LEVELS ARE FOUND IN EMBRYO AND ADULT.
CC       LEVELS DECREASE DURING THE FIRST AND SECOND LARVAL INSTAR AND THEN
CC       DECREASE IN THIRD INSTAR LARVAE AND EARLY PUPAE.
CC   -!- SIMILARITY: Belongs to the ZW10 family.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-44, MET-81 OR MET-100
CC       IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64390; CAB76122.1; -.
DR   EMBL; AE003424; AAF45794.1; -.
DR   EMBL; AL138972; CAB72295.1; -.
DR   EMBL; AL121804; CAB72295.1; JOINED.
DR   EMBL; AL121804; CAB65854.1; -.
DR   PIR; A43275; A43275.
DR   FlyBase; FBgn0004643; mit(1)15.
DR   GO; GO:0005828; C:kinetochore microtubule; IDA.
DR   GO; GO:0005699; C:kinetochore; IDA.
DR   GO; GO:0000070; P:mitotic chromosome segregation; IMP.
KW   Cell cycle; Meiosis; Mitosis; Nuclear protein; Centromere.
FT   CONFLICT     58     58       L -> M (IN REF. 1).
FT   CONFLICT    203    203       D -> A (IN REF. 1).
FT   CONFLICT    248    248       C -> A (IN REF. 3).
FT   CONFLICT    293    294       HV -> QL (IN REF. 2).
FT   CONFLICT    366    366       V -> A (IN REF. 1).
FT   CONFLICT    626    626       D -> H (IN REF. 3).
SQ   SEQUENCE   721 AA;  82263 MW;  1341BC2BF752188D CRC64;

  Query Match             65.5%;  Score 36;  DB 1;  Length 721;
  Best Local Similarity   66.7%;  Pred. No. 65;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTL 9
              :|:|||| |
Db        589 LKNVWANVL 597


RESULT 8
METE_CATRO
ID   METE_CATRO     STANDARD;      PRT;   765 AA.
AC   Q42699;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
DE   (EC 2.1.1.14) (Vitamin-B12-independent methionine synthase isozyme)
DE   (Cobalamin-independent methionine synthase isozyme).
GN   METE.
OS   Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Gentianales; Apocynaceae; Rauvolfioideae; Vinceae;
OC   Catharanthus.
OX   NCBI_TaxID=4058;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95324563; PubMed=7601135;
RA   Eichel J., Gonzalez J.C., Hotze M., Matthews R.G., Schroeder J.;
RT   "Vitamin-B12-independent methionine synthase from a higher plant
RT   (Catharanthus roseus). Molecular characterization, regulation,
RT   heterologous expression, and enzyme properties.";
RL   Eur. J. Biochem. 230:1053-1058(1995).
CC   -!- FUNCTION: Catalyzes the transfer of a methyl group from 5-
CC       methyltetrahydrofolate to homocysteine resulting in methionine
CC       formation.
CC   -!- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltri-L-glutamate + L-
CC       homocysteine = tetrahydropteroyltri-L-glutamate + L-methionine.
CC   -!- COFACTOR: Zinc (By similarity).
CC   -!- PATHWAY: Terminal step in the de novo biosynthesis of methionine.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- MISCELLANEOUS: HAS AN ABSOLUTE REQUIREMENT FOR A POLYGLUTAMYLATED
CC       FOLATE AS SUBSTRATE.
CC   -!- SIMILARITY: Belongs to the vitamin-B12 independent methionine
CC       synthase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83499; CAA58474.1; -.
DR   PIR; S57636; S57636.
DR   InterPro; IPR006276; Met_syn_B12ind.
DR   InterPro; IPR002629; Methionine_synth.
DR   Pfam; PF01717; Methionine_synt; 1.
DR   ProDom; PD004692; Methionine_synt; 2.
```

```
R   TIGRFAMs; TIGR01371; met_syn_B12ind; 1.
W   Transferase; Methyltransferase; Methionine biosynthesis; Zinc.
T   METAL       647    647       ZINC (BY SIMILARITY).
T   METAL       649    649       ZINC (BY SIMILARITY).
T   METAL       733    733       ZINC (BY SIMILARITY).
Q   SEQUENCE   765 AA;  84856 MW;  7398A1DC8AE960CE CRC64;

 Query Match             65.5%;  Score 36;  DB 1;  Length 765;
 Best Local Similarity   38.5%;  Pred. No. 69;
 Matches    5;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

/          1 XKXVWANTLKAAX 13
             ::::||| | |::
>        295 GRNIWANDLAASL 307


ESULT 9
ETE_MESCR
D   METE_MESCR     STANDARD;      PRT;   765 AA.
C   P93263;
T   30-MAY-2000 (Rel. 39, Created)
T   30-MAY-2000 (Rel. 39, Last sequence update)
T   16-OCT-2001 (Rel. 40, Last annotation update)
E   5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
E   (EC 2.1.1.14) (Vitamin-B12-independent methionine synthase isozyme)
E   (Cobalamin-independent methionine synthase isozyme).
N   METE.
S   Mesembryanthemum crystallinum (Common ice plant).
C   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
C   Caryophyllales; Aizoaceae; Mesembryanthemum.
X   NCBI_TaxID=3544;
N   [1]
P   SEQUENCE FROM N.A.
A   Michalowski C.B., Quigley-Landreau F., Bohnert H.J.;
T   "Mesembryanthemum crystallinum methionine synthase mRNA.";
L   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
C   -!- FUNCTION: Catalyzes the transfer of a methyl group from 5-
C       methyltetrahydrofolate to homocysteine resulting in methionine
C       formation (By similarity).
C   -!- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltri-L-glutamate + L-
C       homocysteine = tetrahydropteroyltri-L-glutamate + L-methionine.
C   -!- COFACTOR: Zinc (By similarity).
C   -!- PATHWAY: Terminal step in the de novo biosynthesis of methionine.
C   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
C   -!- SIMILARITY: Belongs to the vitamin-B12 independent methionine
C       synthase family.
C   ------------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ------------------------------------------------------------------------
R   EMBL; U84889; AAB41896.1; -.
R   PIR; T12575; T12575.
R   InterPro; IPR006276; Met_syn_B12ind.
R   InterPro; IPR002629; Methionine_synth.
R   Pfam; PF01717; Methionine_synt; 1.
R   ProDom; PD004692; Methionine_synt; 2.
R   TIGRFAMs; TIGR01371; met_syn_B12ind; 1.
W   Transferase; Methyltransferase; Methionine biosynthesis; Zinc.
T   METAL       647    647       ZINC (BY SIMILARITY).
T   METAL       649    649       ZINC (BY SIMILARITY).
T   METAL       733    733       ZINC (BY SIMILARITY).
Q   SEQUENCE   765 AA;  84821 MW;  8C510684202EA000 CRC64;

 Query Match             65.5%;  Score 36;  DB 1;  Length 765;
 Best Local Similarity   38.5%;  Pred. No. 69;
 Matches    5;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              ::::||| | |::
Db        295 GRNIWANDLAASL 307


RESULT 10
SIR4_YEAST
ID   SIR4_YEAST     STANDARD;      PRT;  1358 AA.
AC   P11978;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Regulatory protein SIR4 (Silent information regulator 4).
GN   SIR4 OR STE9 OR ASD1 OR UTH2 OR YDR227W OR YD9934.12.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88142836; PubMed=3325825;
RA   Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT   "Functional domains of SIR4, a gene required for position effect
RT   regulation in Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:4441-4452(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95192063; PubMed=7885847;
RA   Davies C.J., Hutchison C.A. III;
RT   "Insertion site specificity of the transposon Tn3.";
RL   Nucleic Acids Res. 23:507-514(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVIEW.
RX   MEDLINE=21579783; PubMed=11722841;
RA   Gasser S.M., Cockell M.M.;
RT   "The molecular biology of the SIR proteins.";
RL   Gene 279:1-16(2001).
CC   -!- FUNCTION: The proteins SIR1 through SIR4 are required for
CC       transcriptional repression of the silent mating type loci, HML and
CC       HMR. The proteins SIR2 through SIR4 repress mulitple loci by
CC       modulating chromatin structure. Involves the compaction of
CC       chromatin fiber into a more condensed form.
CC   -!- SUBUNIT: Homodimer and interacts with SIR1, SIR2, SIR3 and RAP1 C-
CC       terminus.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M37249; AAA20881.1; -.
DR   EMBL; U13239; AAC33144.1; -.
DR   EMBL; Z48612; CAA88507.1; -.
DR   PIR; A29360; A29360.
DR   GermOnline; 140719; -.
DR   SGD; S0002635; SIR4.
DR   GO; GO:0000783; C:nuclear telomere cap complex; IDA.
DR   GO; GO:0005724; C:nuclear telomeric heterochromatin; IDA.
DR   GO; GO:0006303; P:double-strand break repair via nonhomologou. . .; IDA.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW   Coiled coil.
FT   DOMAIN     1277   1347       COILED COIL (POTENTIAL).
FT   VARIANT     994    994       P -> L.
SQ   SEQUENCE   1358 AA;  152061 MW;  9C698765964F094E CRC64;
```

```
 Query Match             65.5%;  Score 36;  DB 1;  Length 1358;
 Best Local Similarity   46.2%;  Pred. No. 1.3e+02;
 Matches    6;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|:::|| ||:|:
Db       1012 LKDIFANNLKSAI 1024


ESULT 11
322_HAEIN
D   Y322_HAEIN     STANDARD;      PRT;   134 AA.
C   Q57122; O05016;
T   01-NOV-1997 (Rel. 35, Created)
T   01-NOV-1997 (Rel. 35, Last sequence update)
T   16-OCT-2001 (Rel. 40, Last annotation update)
D   Hypothetical protein HI0322.
G   HI0322.
S   Haemophilus influenzae.
C   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
C   Pasteurellaceae; Haemophilus.
X   NCBI_TaxID=727;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=Rd / KW20 / ATCC 51907;
X   MEDLINE=95350630; PubMed=7542800;
A   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
A   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
A   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
A   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
A   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
A   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
A   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
A   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
A   Venter J.C.;
T   "Whole-genome random sequencing and assembly of Haemophilus influenzae
T   Rd.";
J   Science 269:496-512(1995).
C   -!- SIMILARITY: TO H.INFLUENZAE HI0947.
C   -!- SIMILARITY: TO B.NODOSUS VIRULENCE-ASSOCIATED PROTEIN C
C       (VAPC).
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; U32717; AAC21985.1; -.
R   PIR; G64061; G64061.
R   TIGR; HI0322; -.
R   InterPro; IPR002716; PIN.
R   InterPro; IPR006596; PINc.
R   Pfam; PF01850; PIN; 1.
R   SMART; SM00670; PINc; 1.
K   Hypothetical protein; Complete proteome.
S   SEQUENCE   134 AA;  15726 MW;  7EC5014217A854F9 CRC64;

 Query Match             63.6%;  Score 35;  DB 1;  Length 134;
 Best Local Similarity   100.0%;  Pred. No. 17;
 Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 WANTLK 10
              ||||||
Db         84 WANTLK 89


ESULT 12
272_AQUAE
D   Y272_AQUAE     STANDARD;      PRT;   149 AA.
AC   O66629;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AQ_272.
GN   AQ_272.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000681; AAC06587.1; -.
DR   PIR; A70325; A70325.
KW   Hypothetical protein; Coiled coil; Complete proteome.
FT   DOMAIN      111    140       COILED COIL (POTENTIAL).
SQ   SEQUENCE   149 AA;  17945 MW;  92E2623B513E79E3 CRC64;

  Query Match             63.6%;  Score 35;  DB 1;  Length 149;
  Best Local Similarity   60.0%;  Pred. No. 19;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLK 10
              :|: | ||||
Db         42 PKEEWENTLK 51


RESULT 13
YP1_AGRT4
ID   YP1_AGRT4      STANDARD;      PRT;   172 AA.
AC   P04028;
DT   23-OCT-1986 (Rel. 02, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein 1 (Gene 5 protein).
OS   Agrobacterium tumefaciens (strain Ach5).
OG   Plasmid pTiAch5.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176298;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93225814; PubMed=8469115;
RA   Turk S.C.H.J., Nester E.W., Hooykaas P.J.J.;
RT   "The virA promoter is a host-range determinant in Agrobacterium
RT   tumefaciens.";
RL   Mol. Microbiol. 7:719-724(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94035196; PubMed=8220492;
RA   Guevara-Garcia A., Mosqueda-Cano G., Argueello-Astorga G.,
RA   Simpson J., Herrera-Estrella L.;
RT   "Tissue-specific and wound-inducible pattern of expression of the
RT   mannopine synthase promoter is determined by the interaction between
RT   positive and negative cis-regulatory elements.";
RL   Plant J. 4:495-505(1993).
RN   [3]
```

```
P   SEQUENCE FROM N.A.
A   Barker R.F., Idler K.B., Thompson D.V., Kemp J.D.;
T   "Nucleotide sequence of the T-DNA region from the Agrobacterium
T   tumefaciens octopine Ti plasmid pTi15955.";
L   Plant Mol. Biol. 2:335-350(1983).
N   [4]
P   PRELIMINARY SEQUENCE FROM N.A.
X   MEDLINE=84207942; PubMed=6327292;
A   Gielen J., de Beuckeleer M., Seurinck J., Deboeck F., de Greve H.,
A   Lemmers M., van Montagu M., Schell J.;
T   "The complete nucleotide sequence of the TL-DNA of the Agrobacterium
T   tumefaciens plasmid pTiAch5.";
L   EMBO J. 3:835-846(1984).
N   [5]
P   SEQUENCE FROM N.A.
A   Winans S.C., Zhu J., Oger P.M., Schrammeijer B., Hooykaas P.J.,
A   Farrand S.K.;
T   "Octopine-type Ti plasmid sequence.";
L   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
C   ------------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ------------------------------------------------------------------------
R   EMBL; X00493; CAA25163.1; -.
R   EMBL; AF242881; AAF77120.1; -.
R   PIR; A04496; QQAG1T.
R   PIR; S28683; S28683.
R   InterPro; IPR006064; Glycosidase.
R   Pfam; PF02027; RolB_RolC; 1.
W   Crown gall tumor; Plasmid; Hypothetical protein.
Q   SEQUENCE   172 AA;  19830 MW;  956C85F4B0A96D88 CRC64;

 Query Match             63.6%;  Score 35;  DB 1;  Length 172;
 Best Local Similarity   60.0%;  Pred. No. 22;
 Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

y          1 XKXVWANTLK 10
             :|:|| || |
b        139 QKKVWQNTSK 148


ESULT 14
GE2_HUMAN
D   CGE2_HUMAN     STANDARD;      PRT;   404 AA.
C   O96020; O95439;
T   15-JUL-1999 (Rel. 38, Created)
T   15-JUL-1999 (Rel. 38, Last sequence update)
T   15-MAR-2004 (Rel. 43, Last annotation update)
E   G1/S-Specific cyclin E2.
N   CCNE2.
S   Homo sapiens (Human).
C   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X   NCBI_TaxID=9606;
N   [1]
P   SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
C   TISSUE=Fetal lung;
X   MEDLINE=99077999; PubMed=9858585;
A   Gudas J.M., Payton M., Thukral S., Chen E., Bass M., Robinson M.O.,
A   Coats S.;
T   "Cyclin E2, a novel G1 cyclin that binds Cdk2 and is aberrantly
T   expressed in human cancers.";
L   Mol. Cell. Biol. 19:612-622(1999).
N   [2]
P   SEQUENCE FROM N.A.
C   TISSUE=B-cell;
X   MEDLINE=99054662; PubMed=9840927;
A   Lauper N., Beck A.R.P., Cariou S., Richman L., Hofmann K., Reith W.,
RA  Slingerland J.M., Amati B.;
RT  "Cyclin E2: a novel CDK2 partner in the late G1 and S phases of the
RT  mammalian cell cycle.";
RL  Oncogene 17:2637-2643(1998).
RN  [3]
RP  SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND MUTAGENESIS OF THR-392.
RC  TISSUE=Keratinocytes;
RX  MEDLINE=99054678; PubMed=9840943;
RA  Zariwala M., Liu J., Xiong Y.;
RT  "Cyclin E2, a novel human G1 cyclin and activating partner of CDK2 and
RT  CDK3, is induced by viral oncoproteins.";
RL  Oncogene 17:2787-2798(1998).
CC  -!- FUNCTION: Essential for the control of the cell cycle at the late
CC      G1 and early S phase.
CC  -!- SUBUNIT: Interacts with the CDK2 (in vivo) and CDK3 (in vitro)
CC      protein kinases to form a serine/threonine kinase holoenzyme
CC      complex. The cyclin subunit imparts substrate specificity to the
CC      complex.
CC  -!- SUBCELLULAR LOCATION: Nuclear.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=Long;
CC        IsoId=O96020-1; Sequence=Displayed;
CC      Name=Short; Synonyms=SV;
CC        IsoId=O96020-2; Sequence=VSP_001256;
CC  -!- TISSUE SPECIFICITY: According to Ref.1: highest levels in adult
CC      testis, thymus and brain. Lower levels in placenta, spleen and
CC      colon. Consistently elevated levels in tumor-derived cells
CC      compared to nontransformed proliferating cells. According to
CC      Ref.2: low levels in thymus, prostate, brain, skeletal muscle, and
CC      kidney. Elevated levels in lung. According to Ref.3: highly
CC      expressed in testis, placenta, thymus and brain. In a lesser
CC      extent in small intestine and colon.
CC  -!- INDUCTION: Activated by papilloma viral oncoproteins E6 and E7
CC      which bind to and inactivate p53 and Rb, respectively.
CC  -!- PTM: PHOSPHORYLATION BY CDK2 TRIGGERS ITS RELEASE FROM CDK2 AND
CC      DEGRADATION VIA THE UBIQUITIN PROTEASOME PATHWAY (BY SIMILARITY).
CC  -!- SIMILARITY: Belongs to the cyclin family. Cyclin E subfamily.
CC  ------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ------------------------------------------------------------------------
DR  EMBL; AF106690; AAD08816.1; -.
DR  EMBL; AF112857; AAD08819.1; -.
DR  EMBL; AF091433; AAC80528.1; -.
DR  EMBL; AF102778; AAC78145.1; -.
DR  Genew; HGNC:1590; CCNE2.
DR  GK; O96020; -.
DR  MIM; 603775; -.
DR  GO; GO:0000075; P:cell cycle checkpoint; TAS.
DR  GO; GO:0000079; P:regulation of CDK activity; TAS.
DR  InterPro; IPR006670; Cyclin.
DR  InterPro; IPR004367; Cyclin_Cterm.
DR  InterPro; IPR006671; Cyclin_N.
DR  Pfam; PF00134; cyclin; 1.
DR  Pfam; PF02984; cyclin_C; 1.
DR  SMART; SM00385; CYCLIN; 1.
DR  PROSITE; PS00292; CYCLINS; 1.
KW  Cyclin; Cell cycle; Cell division; Phosphorylation;
KW  Alternative splicing; Nuclear protein.
FT  MOD_RES     392    392       PHOSPHORYLATION (PROBABLE).
FT  VARSPLIC    167    211       Missing (in isoform Short).
FT                               /FTId=VSP_001256.
FT  MUTAGEN     392    392       T->A: INCREASE OF STEADY STATE LEVEL.
SQ  SEQUENCE   404 AA;  46757 MW;  D7DC9BEEF3FD62EC CRC64;

  Query Match             63.6%;  Score 35;  DB 1;  Length 404;
  Best Local Similarity   60.0%;  Pred. No. 54;
```

```
Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

          1 XKXVWANTLK 10
            :|:|| | ||
        110 SKEVWLNMLK 119


SULT 15
YA_PYRAB
    AMYA_PYRAB     STANDARD;      PRT;   655 AA.
    Q9V298;
    16-OCT-2001 (Rel. 40, Created)
    16-OCT-2001 (Rel. 40, Last sequence update)
    10-OCT-2003 (Rel. 42, Last annotation update)
    Alpha-amylase (EC 3.2.1.1).
    AMYA OR PYRAB01760 OR PAB0118.
    Pyrococcus abyssi.
    Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
    Pyrococcus.
    NCBI_TaxID=29292;
    [1]
    SEQUENCE FROM N.A.
    STRAIN=GE5 / Orsay;
    MEDLINE=22511545; PubMed=12622808;
    Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
    Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
    Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
    "An integrated analysis of the genome of the hyperthermophilic
    archaeon Pyrococcus abyssi.";
    Mol. Microbiol. 47:1495-1512(2003).
    -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
        linkages in oligosaccharides and polysaccharides.
    -!- PATHWAY: Polysaccharide degradation.
    -!- SIMILARITY: Belongs to family 57 of glycosyl hydrolases.
    ----------------------------------------------------------------------
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    ----------------------------------------------------------------------
    EMBL; AJ248283; CAB49100.1; -.
    PIR; E75206; E75206.
    InterPro; IPR004300; Glyco_hydro_57.
    Pfam; PF03065; Glyco_hydro_57; 1.
    Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
    SEQUENCE   655 AA;  77296 MW;  7F6F920B1A00EECE CRC64;

Query Match             63.6%;  Score 35;  DB 1;  Length 655;
Best Local Similarity   36.4%;  Pred. No. 90;
Matches    4;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

          1 XKXVWANTLKA 11
            ::::|:| :||
        370 RRAIWSNLIKA 380


arch completed: May 17, 2004, 13:50:02
b time : 11 secs
```

GenCore version 5.1.6

1 protein - protein search, using sw model

un on: May 17, 2004, 13:39:49 ; Search time 39 Seconds
(without alignments)
105.173 Million cell updates/sec

itle: US-09-458-299A-4226
erfect score: 55
equence: 1 XKXVWANTLKAAX 13

coring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

earched: 1017041 seqs, 315518202 residues

otal number of hits satisfying chosen parameters: 1017041

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	44	80.0	333	16	Q88Y31	Q88y31 lactobacill
2	41	74.5	3322	16	Q8XQZ5	Q8xqz5 ralstonia s
3	40	72.7	537	2	Q54410	Q54410 streptomyce
4	40	72.7	541	16	Q9FCD7	Q9fcd7 streptomyce
5	40	72.7	543	16	Q8NQ84	Q8nq84 corynebacte
6	40	72.7	543	16	Q8FT92	Q8ft92 corynebacte
7	40	72.7	1695	5	O62604	O62604 polyorchis
8	39	70.9	111	16	Q9RZ50	Q9rz50 deinococcus
9	39	70.9	150	15	O89784	O89784 human immun
10	39	70.9	202	15	Q9E4D3	Q9e4d3 human immun
11	39	70.9	217	16	Q8UJM1	Q8ujm1 agrobacteri
12	39	70.9	297	10	Q9AX92	Q9ax92 oryza sativ
13	39	70.9	318	5	Q9BJ36	Q9bj36 toxoplasma
14	39	70.9	328	15	Q8QF59	Q8qf59 human immun
15	39	70.9	423	10	Q8LDZ9	Q8ldz9 arabidopsis
16	39	70.9	490	16	Q7V4I3	Q7v4i3 prochloroco
17	38	69.1	108	15	O55891	O55891 human immun
18	38	69.1	171	16	Q7WDL6	Q7wdl6 bordetella
19	38	69.1	171	16	Q7W2M0	Q7w2m0 bordetella
20	38	69.1	196	16	Q81RH1	Q81rh1 bacillus an
21	38	69.1	196	16	Q81EC3	Q81ec3 bacillus ce
22	38	69.1	246	16	Q8UFM6	Q8ufm6 agrobacteri
23	38	69.1	275	16	Q8DDP6	Q8ddp6 vibrio vuln
24	38	69.1	402	12	Q9J2L3	Q9j2l3 macaca mula
25	38	69.1	402	12	Q9WRS9	Q9wrs9 macaca mula
26	38	69.1	446	16	Q7VQ80	Q7vq80 chlamydia p
27	38	69.1	485	16	Q9Z957	Q9z957 chlamydia p
28	38	69.1	485	16	Q9JRX6	Q9jrx6 chlamydia p
29	38	69.1	538	3	O43011	O43011 schizosacch
30	38	69.1	937	9	Q9G097	Q9g097 lactococcus
31	38	69.1	1025	9	Q9AYV8	Q9ayv8 lactococcus
32	38	69.1	1628	16	Q98QZ7	Q98qz7 mycoplasma
33	37.5	68.2	179	8	Q94NU7	Q94nu7 phoxinus er
34	37.5	68.2	611	8	Q951J2	Q951j2 phoxinus er
35	37	67.3	77	15	P89371	P89371 human immun
36	37	67.3	103	15	Q69936	Q69936 human immun
37	37	67.3	168	15	Q7ZFV8	Q7zfv8 human immun
38	37	67.3	204	15	Q9Q445	Q9q445 human immun
39	37	67.3	208	15	Q9YV88	Q9yv88 human immun
40	37	67.3	211	15	Q9YV87	Q9yv87 human immun
41	37	67.3	211	15	O90028	O90028 human immun
42	37	67.3	211	15	Q9YV89	Q9yv89 human immun
43	37	67.3	220	15	Q73039	Q73039 human immun
44	37	67.3	226	15	Q9QL73	Q9ql73 human immun
45	37	67.3	233	16	Q7UN93	Q7un93 rhodopirell

ALIGNMENTS

RESULT 1
Q88Y31
ID Q88Y31 PRELIMINARY; PRT; 333 AA.
AC Q88Y31;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN LP_0967.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935254; CAD63526.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 333 AA; 37801 MW; C9175ACD4E2EDA0A CRC64;

Query Match 80.0%; Score 44; DB 16; Length 333;
Best Local Similarity 63.6%; Pred. No. 9.9;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy          3 XVWANTLKAAX 13
              ::| ||||||:
Db        125 WIWVNTLKAAQ 135
```

RESULT 2
Q8XQZ5

```
D   Q8XQZ5      PRELIMINARY;      PRT;  3322 AA.
C   Q8XQZ5;
T   01-MAR-2002 (TrEMBLrel. 20, Created)
T   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
T   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E   Probable hemagglutinin-related protein.
N   RSP1073 OR RS02477.
S   Ralstonia solanacearum (Pseudomonas solanacearum).
G   Plasmid megaplasmid.
C   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
C   Burkholderiaceae; Ralstonia.
X   NCBI_TaxID=305;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=GMI1000;
X   MEDLINE=21681879; PubMed=11823852;
A   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
A   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
A   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
A   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
A   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
A   Weissenbach J., Boucher C.A.;
T   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
L   Nature 415:497-502(2002).
R   EMBL; AL646082; CAD18224.1; -.
R   GO; GO:0046821; C:extrachromosomal DNA; IEA.
R   InterPro; IPR008619; Fil_haemagg.
R   InterPro; IPR008638; Haemagg_act.
R   Pfam; PF05594; Fil_haemagg; 19.
R   Pfam; PF05860; Haemagg_act; 1.
W   Plasmid; Complete proteome.
Q   SEQUENCE   3322 AA;  332591 MW;  47120FE79BF8450C CRC64;

 Query Match             74.5%;  Score 41;  DB 16;  Length 3322;
 Best Local Similarity   54.5%;  Pred. No. 4.2e+02;
 Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

/          3 XVWANTLKAAX 13
             ::|||||| ::
>        245 GIWANTLKVSA 255


ESULT 3
54410
D   Q54410      PRELIMINARY;      PRT;   537 AA.
C   Q54410;
T   01-NOV-1996 (TrEMBLrel. 01, Created)
T   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
T   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E   Tripeptidylaminopeptidase precursor.
S   Streptomyces lividans.
C   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
C   Streptomycineae; Streptomycetaceae; Streptomyces.
X   NCBI_TaxID=1916;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=96090168; PubMed=7487044;
A   Butler M.J., Binnie C., DiZonno M.A., Krygsman P., Soltes G.A.,
A   Soostmeyer G., Walczyk E., Malek L.T.;
T   "Cloning and characterization of a gene encoding a secreted
T   tripeptidyl aminopeptidase from Streptomyces lividans 66.";
L   Appl. Environ. Microbiol. 61:3145-3150(1995).
R   EMBL; L27466; AAA92338.1; -.
R   MEROPS; S33.002; -.
R   GO; GO:0004177; F:aminopeptidase activity; IEA.
R   GO; GO:0003824; F:catalytic activity; IEA.
R   InterPro; IPR000073; A/b_hydrolase.
R   InterPro; IPR008262; Lipase_AS.
R   Pfam; PF00561; abhydrolase; 1.
R   PROSITE; PS00120; LIPASE_SER; 1.
W   Aminopeptidase; Signal.
T   SIGNAL        1     39       POTENTIAL.
T   CHAIN        40    537       TRIPEPTIDYLAMINOPEPTIDASE.
SQ   SEQUENCE   537 AA;  58273 MW;  850703374BC4DEC9 CRC64;

  Query Match             72.7%;  Score 40;  DB 2;  Length 537;
  Best Local Similarity   63.6%;  Pred. No. 90;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKA 11
              : :||||| ||
Db        142 KSAVWANTAKA 152


RESULT 4
Q9FCD7
ID   Q9FCD7      PRELIMINARY;      PRT;   541 AA.
AC   Q9FCD7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative secreted tripeptidylaminopeptidase.
GN   SCO1230 OR 2SCG1.05C.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Brown S.P., Harris D.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL939108; CAC01454.1; -.
DR   MEROPS; S33.002; -.
DR   GO; GO:0004177; F:aminopeptidase activity; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   InterPro; IPR000073; A/b_hydrolase.
DR   InterPro; IPR008262; Lipase_AS.
DR   Pfam; PF00561; abhydrolase; 1.
DR   PROSITE; PS00120; LIPASE_SER; 1.
KW   Aminopeptidase; Complete proteome.
SQ   SEQUENCE   541 AA;  58535 MW;  01BA2E6F70B124DB CRC64;

  Query Match             72.7%;  Score 40;  DB 16;  Length 541;
  Best Local Similarity   63.6%;  Pred. No. 91;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
          1 XKXVWANTLKA 11
            : :||||| ||
        146 KSAVWANTAKA 156


SULT 5
NQ84
   Q8NQ84      PRELIMINARY;      PRT;   543 AA.
   Q8NQ84;
   01-OCT-2002 (TrEMBLrel. 22, Created)
   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
   ABC-type transporter, duplicated ATPase component.
   CGL1554.
   Corynebacterium glutamicum (Brevibacterium flavum).
   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
   NCBI_TaxID=1718;
   [1]
   SEQUENCE FROM N.A.
   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
   Nakagawa S.;
   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
   EMBL; AP005278; BAB98947.1; -.
   GO; GO:0016020; C:membrane; IEA.
   GO; GO:0005524; F:ATP binding; IEA.
   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
   GO; GO:0006810; P:transport; IEA.
   InterPro; IPR003439; ABC_transporter.
   Pfam; PF00005; ABC_tran; 2.
   ProDom; PD000006; ABC_transporter; 2.
   PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
   PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
   Complete proteome.
   SEQUENCE   543 AA;  58866 MW;  E5062AE37A4DDEF1 CRC64;

Query Match             72.7%;  Score 40;  DB 16;  Length 543;
Best Local Similarity   46.2%;  Pred. No. 91;
Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

          1 XKXVWANTLKAAX 13
            :|:|| ||::| :
        414 DKSVWQNTIEACA 426


SULT 6
FT92
   Q8FT92      PRELIMINARY;      PRT;   543 AA.
   Q8FT92;
   01-MAR-2003 (TrEMBLrel. 23, Created)
   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
   Putative ABC transporter ATP-binding protein.
   CE1678.
   Corynebacterium efficiens.
   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
   NCBI_TaxID=152794;
   [1]
   SEQUENCE FROM N.A.
   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
   Usuda Y., Sugimoto S.;
   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
   EMBL; AP005219; BAC18488.1; -.
   GO; GO:0016020; C:membrane; IEA.
   GO; GO:0005524; F:ATP binding; IEA.
   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
   GO; GO:0000166; F:nucleotide binding; IEA.
   GO; GO:0006810; P:transport; IEA.
   InterPro; IPR003593; AAA_ATPase.
   InterPro; IPR003439; ABC_transporter.
   Pfam; PF00005; ABC_tran; 2.
   ProDom; PD000006; ABC_transporter; 2.
   SMART; SM00382; AAA; 2.
   PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
   PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
   ATP-binding; Complete proteome.
   SEQUENCE   543 AA;  58713 MW;  5C29C5E8A02ED77A CRC64;

  Query Match             72.7%;  Score 40;  DB 16;  Length 543;
  Best Local Similarity   46.2%;  Pred. No. 91;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|:|| ||::| :
Db        414 DKSVWQNTIEACA 426


RESULT 7
O62604
ID   O62604      PRELIMINARY;      PRT;  1695 AA.
AC   O62604;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Voltage-gated sodium channel alpha subunit.
GN   SCN1.
OS   Polyorchis penicillatus (Hydromedusa).
OC   Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC   Polyorchidae; Polyorchis.
OX   NCBI_TaxID=6091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98205797; PubMed=9535741;
RA   Spafford J.D., Spencer A.N., Gallin W.J.;
RT   "A putative voltage-gated sodium channel alpha subunit (PpSCN1) from
RT   the hydrozoan jellyfish, Polyorchis penicillatus: structural
RT   comparisons and evolutionary considerations.";
RL   Biochem. Biophys. Res. Commun. 244:772-780(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Spafford J.D., Spencer A.N., Gallin W.J.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF047380; AAC38974.1; -.
DR   EMBL; AF047379; AAC09306.1; -.
DR   PIR; JE0084; JE0084.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001518; C:voltage-gated sodium channel complex; IEA.
DR   GO; GO:0005261; F:cation channel activity; IEA.
DR   GO; GO:0005248; F:voltage-gated sodium channel activity; IEA.
DR   GO; GO:0006812; P:cation transport; IEA.
DR   GO; GO:0006814; P:sodium ion transport; IEA.
DR   InterPro; IPR001682; Ca/Na_pore.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR005820; M+channel_nlg.
DR   InterPro; IPR001696; Na_channel.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00170; NACHANNEL.
KW   Ionic channel; Transmembrane.
SQ   SEQUENCE   1695 AA;  194545 MW;  FDA141CFB913BEA1 CRC64;

  Query Match             72.7%;  Score 40;  DB 5;  Length 1695;
  Best Local Similarity   77.8%;  Pred. No. 3.1e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 WANTLKAAX 13
              | |||||:
Db       1192 WVNTLKAAS 1200
```

```
ESULT 8
)9RZ50
D   Q9RZ50      PRELIMINARY;      PRT;   111 AA.
AC  Q9RZ50;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Hypothetical protein DRA0104.
GN  DRA0104.
OS  Deinococcus radiodurans.
OC  Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC  Deinococcaceae; Deinococcus.
OX  NCBI_TaxID=1299;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX  MEDLINE=20036896; PubMed=10567266;
RA  White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA  Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA  Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA  Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA  Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA  Fraser C.M.;
RT  "Genome sequence of the radioresistant bacterium Deinococcus
RT  radiodurans R1.";
RL  Science 286:1571-1577(1999).
DR  EMBL; AE001862; AAF12331.1; -.
DR  PIR; B75605; B75605.
DR  TIGR; DRA0104; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE   111 AA;  12494 MW;  CC3DB4CCAE32870B CRC64;

 Query Match             70.9%;  Score 39;  DB 16;  Length 111;
 Best Local Similarity   63.6%;  Pred. No. 25;
 Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 XVWANTLKAAX 13
              :|||:| ||:
Db         62 SVWANSLDAAI 72


ESULT 9
89784
ID  O89784      PRELIMINARY;      PRT;   150 AA.
AC  O89784;
DT  01-NOV-1998 (TrEMBLrel. 08, Created)
DT  01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Envelope glycoprotein (Fragment).
GN  ENV.
OS  Human immunodeficiency virus 1.
OC  Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX  NCBI_TaxID=11676;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=GR30;
RX  MEDLINE=98278642; PubMed=9618080;
RA  Nasioulas G., Paraskevis D., Paparizos V., Lazanas M.,
RA  Karafoulidou A., Hatzakis A.;
RT  "Genotypic characterization of human immunodeficiency virus type 1 in
RT  Greece. Multicentre Study on HIV-1 Heterogeneity.";
RL  AIDS Res. Hum. Retroviruses 14:685-690(1998).
DR  EMBL; AF049305; AAD05099.1; -.
DR  GO; GO:0019028; C:viral capsid; IEA.
DR  GO; GO:0019031; C:viral envelope; IEA.
DR  GO; GO:0005198; F:structural molecule activity; IEA.
DR  InterPro; IPR000777; GP120.
DR  Pfam; PF00516; GP120; 1.
KW  AIDS; Coat protein; Glycoprotein.
FT  NON_TER       1      1
FT  NON_TER     150    150
SQ  SEQUENCE   150 AA;  16573 MW;  4F07A85095921303 CRC64;

  Query Match             70.9%;  Score 39;  DB 15;  Length 150;
  Best Local Similarity   60.0%;  Pred. No. 34;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLK 10
              :|: |:||||
Db        110 SKAAWSNTLK 119


RESULT 10
Q9E4D3
ID  Q9E4D3      PRELIMINARY;      PRT;   202 AA.
AC  Q9E4D3;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Envelope glycoprotein (Fragment).
GN  ENV.
OS  Human immunodeficiency virus 1.
OC  Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX  NCBI_TaxID=11676;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Leroux C.;
RT  "HIV-1 shedding.";
RL  Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF256452; AAG16113.1; -.
DR  GO; GO:0019028; C:viral capsid; IEA.
DR  GO; GO:0019031; C:viral envelope; IEA.
DR  GO; GO:0005198; F:structural molecule activity; IEA.
DR  InterPro; IPR000777; GP120.
DR  Pfam; PF00516; GP120; 1.
KW  AIDS; Coat protein; Glycoprotein.
FT  NON_TER       1      1
FT  NON_TER     202    202
SQ  SEQUENCE   202 AA;  22620 MW;  EADC7077192B5458 CRC64;

  Query Match             70.9%;  Score 39;  DB 15;  Length 202;
  Best Local Similarity   53.8%;  Pred. No. 47;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|: | |||| |:
Db         67 SKAKWENTLKQAA 79


RESULT 11
Q8UJM1
ID  Q8UJM1      PRELIMINARY;      PRT;   217 AA.
AC  Q8UJM1;
DT  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Hypothetical protein Atu5455.
GN  ATU5455 OR AGR_PAT_668.
OS  Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG  Plasmid AT.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC  Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX  NCBI_TaxID=176299;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21608550; PubMed=11743193;
RA  Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA  Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA  Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA  Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA  Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA  Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA  Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA  Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA  Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
```

```
   Nester E.W.;
   "The genome of the natural genetic engineer Agrobacterium tumefaciens
   C58.";
   Science 294:2317-2323(2001).
   [2]
   SEQUENCE FROM N.A.
   MEDLINE=21608551; PubMed=11743194;
   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
   Cielo C., Slater S.;
   "Genome sequence of the plant pathogen and biotechnology agent
   Agrobacterium tumefaciens C58.";
   Science 294:2323-2328(2001).
   EMBL; AE008966; AAL46142.1; ALT_INIT.
   EMBL; AE007914; AAK90830.1; -.
   PIR; AH3215; AH3215.
   GO; GO:0046821; C:extrachromosomal DNA; IEA.
   Hypothetical protein; Plasmid; Complete proteome.
   SEQUENCE   217 AA;  24711 MW;  0F08878861C15086 CRC64;

Query Match             70.9%;  Score 39;  DB 16;  Length 217;
Best Local Similarity   63.6%;  Pred. No. 51;
Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

          3 XVWANTLKAAX 13
            :||| :||||:
        196 AVWAQSLKAAL 206


SULT 12
AX92
   Q9AX92      PRELIMINARY;      PRT;   297 AA.
   Q9AX92;
   01-JUN-2001 (TrEMBLrel. 17, Created)
   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
   P0501G01.25 protein.
   P0501G01.25.
   Oryza sativa (Rice).
   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
   Ehrhartoideae; Oryzeae; Oryza.
   NCBI_TaxID=4530;
   [1]
   SEQUENCE FROM N.A.
   STRAIN=cv. Nipponbare;
   Sasaki T., Matsumoto T., Yamamoto K.;
   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
   clone:P0501G01.";
   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
   EMBL; AP002819; BAB21096.1; -.
   Gramene; Q9AX92; -.
   SEQUENCE   297 AA;  32599 MW;  CFCA067833D857BD CRC64;

Query Match             70.9%;  Score 39;  DB 10;  Length 297;
Best Local Similarity   46.2%;  Pred. No. 72;
Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

          1 XKXVWANTLKAAX 13
            ::::|| || ||:
        267 GRQIWATTLTAAR 279


SULT 13
BJ36
   Q9BJ36      PRELIMINARY;      PRT;   318 AA.
   Q9BJ36;
   01-JUN-2001 (TrEMBLrel. 17, Created)
   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
   ADP/ATP carrier.
   Toxoplasma gondii.
   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
   Toxoplasma.
   NCBI_TaxID=5811;
   [1]
   SEQUENCE FROM N.A.
   Voncken F., Clayton C.;
   "Mitochondrial ADP/ATP carrier of Toxoplasma gondii.";
   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
   EMBL; AF343580; AAK26384.1; -.
   GO; GO:0005743; C:mitochondrial inner membrane; IEA.
   GO; GO:0005488; F:binding; IEA.
   GO; GO:0006810; P:transport; IEA.
   InterPro; IPR001993; Mitoch_carrier.
   InterPro; IPR002067; Mit_carrier.
   Pfam; PF00153; mito_carr; 3.
   PRINTS; PR00926; MITOCARRIER.
   PROSITE; PS00215; MITOCH_CARRIER; 3.
   SEQUENCE   318 AA;  35242 MW;  AD45301657FDA697 CRC64;

  Query Match             70.9%;  Score 39;  DB 5;  Length 318;
  Best Local Similarity   46.2%;  Pred. No. 78;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLKAAX 13
              :|: ||| |: |:
Db        290 FKGAWANVLRGAG 302


RESULT 14
Q8QF59
ID   Q8QF59      PRELIMINARY;      PRT;   328 AA.
AC   Q8QF59;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Envelope glycoprotein (Fragment).
GN   ENV.
OS   Human immunodeficiency virus 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21624270; PubMed=11752161;
RA   Zhu T., Muthui D., Holte S., Nickle D., Feng F., Brodie S.,
RA   Hwangbo Y., Mullins J.I., Corey L.;
RT   "Evidence for human immunodeficiency virus type 1 replication in vivo
RT   in CD14+ monocytes and its potential role as a source of virus in
RT   patients on highly active antiretroviral therapy.";
RL   J. Virol. 76:707-716(2002).
DR   EMBL; AF405862; AAL76550.1; -.
DR   PIR; A53591; A53591.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR000328; Env_GP41.
DR   InterPro; IPR000777; GP120.
DR   Pfam; PF00516; GP120; 1.
DR   Pfam; PF00517; GP41; 1.
KW   AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     328    328
SQ   SEQUENCE   328 AA;  36035 MW;  D7C60C7D7D7B234F CRC64;

  Query Match             70.9%;  Score 39;  DB 15;  Length 328;
  Best Local Similarity   70.0%;  Pred. No. 80;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 XKXVWANTLK 10
              :|: ||||||
Db        100 SKANWANTLK 109
```

```
SULT 15
LDZ9
)   Q8LDZ9      PRELIMINARY;      PRT;   423 AA.
:   Q8LDZ9;
:   01-OCT-2002 (TrEMBLrel. 22, Created)
:   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
:   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
:   Transcriptional activator RF2a, putative.
:   Arabidopsis thaliana (Mouse-ear cress).
:   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
:   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
:   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
:   NCBI_TaxID=3702;
I   [1]
>   SEQUENCE FROM N.A.
\   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
\   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
:   "Full-length messenger RNA sequences greatly improve genome
:   annotation.";
.   Genome Biol. 0:0-0(2002).
I   [2]
>   SEQUENCE FROM N.A.
\   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
\   Feldmann K.;
:   "Full-Length cDNA from Arabidopsis thaliana.";
.   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
:   EMBL; AY085706; AAM62924.1; -.
:   GO; GO:0005634; C:nucleus; IEA.
:   GO; GO:0003677; F:DNA binding; IEA.
:   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
:   InterPro; IPR004827; TF_bZIP.
:   Pfam; PF00170; bZIP; 1.
:   SMART; SM00338; BRLZ; 1.
:   PROSITE; PS50217; BZIP; 1.
)   SEQUENCE   423 AA;  47062 MW;  FE74A0666FE70B90 CRC64;

 Query Match             70.9%;  Score 39;  DB 10;  Length 423;
 Best Local Similarity   46.2%;  Pred. No. 1.1e+02;
 Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

/           1 XKXVWANTLKAAX 13
              :|::|||:  ||:
>         214 AKRIWANSQSAAR 226


earch completed: May 17, 2004, 13:50:54
ob time : 39 secs
```